# Census 2000
# A.C.E. Methodology
# Volume 1

# Census 2000 A.C.E. Methodology

## Contents

## Volume 1

**Overview** **Tab No.**

## Volume 2

**Sampling**

## Volume 3

**Estimation**

## Volume 4

**1999 Joint Statistical Meetings (JSM) Papers on Census 2000**

## Volume 5

**Quality Assurance**

# ACCURACY AND COVERAGE EVALUATION:

# THEORY AND APPLICATION

Prepared for the February 2 - 3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

Howard Hogan
Decennial Statistical Studies Division
U.S. Bureau of the Census

# The Accuracy and Coverage Evaluation: Theory and Application

## 1. Introduction

The use of the dual system model for human populations is well known in the statistical literature either for its application to measuring the completeness of vital events registration (Chandra Sekar and Deming [1949]; Marks, et al [1974]) or for general use in measuring coverage errors in census data (Marks [1979],Wolter [1986]). Application of the dual system model in the context of the 1990 United States Census, including the issue of census adjustment, is well documented in (Hogan [1992] and Hogan [1993]). In addition, numerous internal and limited circulation documents discuss particular problems, both theoretical and practical. (See for example: U.S. Bureau of the Census [1985].)

What has been lacking is a general discussion of some of the theoretical and practical aspects of the overall design of a dual system estimator in census applications. The Wolter paper, for example, assumes away many of the most difficult practical problems of applying the dual systems model to the actual situations. For example, major issues, such as matching, out-of-scope inclusions in the census and missing data, are all assumed away.

In this paper, we attempt to develop the dual system model for census applications with most, if not all, of its complexity. We seek to show how design decisions fit into a theoretical framework. In addition, we will address several of the issues involved in using the results of the dual system model for census adjustment, especially in those areas where the issues relate directly to the dual system model.

The paper discusses both the general question of designing a post-enumeration survey (PES) using a dual system estimator (DSE), and how these general questions are addressed in the U.S. Census Bureau's plans for the Accuracy and Coverage Evaluation planned as part of Census 2000.[1]

Section 2 reviews the basic dual system theory. Section 3 presents a brief overview of the timing and operations of the Accuracy and Coverage Evaluation (A.C.E.) and notes changes since the 1990 Post-Enumeration Survey (PES). Sections 4, 5 and 6 relate the operations to the theory. Section 7 presents concluding remarks.

[1] 1/Throughout, we will use the terms DSE and PES when a general question is discussed and A.C.E. 2000 when we discuss the specific details of the U. S. 2000 Accuracy and Coverage Evaluation.

## 2. The DSE Model

Following Wolter, let

List A (Census)

| | | In Census | Out of Census | Total |
|---|---|---|---|---|
| | In Survey | $p_{i11}$ | $p_{i12}$ | $p_{i1+}$ |
| List B | Out of Survey | $p_{i21}$ | $p_{i22}$ | $p_{i2+}$ |
| (Survey) | Total | $p_{i+1}$ | $p_{i+2}$ | 1 |

Where $P_{i11}$ denotes the probability that individual "*i*" falls into cell 11, i. e., is captured by both systems. For any class of individuals, there are N people in the true population, then, assuming independence between people (Wolter's autonomous independence), we have in any realization:

| | In Census | Out of Census | Total |
|---|---|---|---|
| In Survey | $N_{11}$ | $N_{12}$ | $N_{1+}$ |
| Out of Survey | $N_{21}$ | $N_{22}$ | $N_{2+}$ |
| Total | $N_{+1}$ | $N_{+2}$ | $N_{++}$ |

where

$N_{11}$ is the number of people counted in both the census and the survey,<br>
$N_{12}$ is the number of people counted in only the survey,<br>
$N_{21}$ is the number of people counted in only the census,<br>
$N_{22}$ is the number of people missed by both the census and the survey,<br>
$N_{1+}$ is the total number of people counted in the survey,<br>
$N_{+1}$ is the total number of people counted in the census, and<br>
$N_{++}$ is the total number of people.

Assuming the capture probabilities of people satisfy $p_{i1+}=p_{1+}$ or $p_{i+1}=p_{+1}$ for all $i=1,\ldots,N$ we have the standard Peterson or Chandrasekaran-Deming model, from which we can estimate

$$\hat{N}_{++} = \frac{N_{+1}\, N_{1+}}{N_{11}} \qquad (1)$$

which is the standard DSE.[2] The DSE can be thought of as:

$$\hat{N}_{++} = N_{+1}\left( {N_{1+}}/{N_{11}} \right) \qquad (2)$$

That is, the total population for the class of people is estimated by the number captured in the census times the inverse ratio of those in both systems to those in the survey (i.e., the inverse of the coverage rate of the Census, as measured by the survey).

All of this is standard. What is not standard is that both the number captured in the Census and the inverse ratio are estimated from, at least in part, a sample survey. Even in the context of a census and a PES, $N_{+1}$ is not the "census count" and $(N_{1+}/N_{11})$ is not the inverse matching rate. Rather, they are constructed measures designed to yield an estimate of the total population.

The DSE will yield a DSE of the population of class *j*, as well as any sum of classes. In the U.S. context, "*j*" might be the household population of a state, of an ethnic group, or perhaps of an ethnic group within a state.

Often, the DSE is combined with a synthetic assumption to produce estimates for areas of geography smaller than that defined by the estimator domain "*j*". Requirements for estimating small or local populations, for example, age by sex, by race, by town, often far exceed the capacity of even a very large E-sample. Using a synthetic assumption, we write

$$CCF_j = \frac{\hat{N}_j}{C_j} \qquad (3)$$

Where,

$CCF_j$ is the net coverage correction factor for group *j*

$\hat{N}_j$ is the DSE of group *j*

[2] Strictly speaking $p_{i1+}$ and $p_{i+1}$ need only be uncorrelated. Zero correlation is most easily approximated (and visualized) if one or both systems have constant capture probabilities.

$C_{jkh}$ is the measure of the population available at the smaller level of geography $k$ (i.e. town, tract, block) and finer demographic subclass $h$.

$$C_j = \sum_k \sum_h C_{jkh}$$

$C_j$ need not equal $N_{+1}$. As we shall see, $N_{+1}$ is the number of people correctly included in the census. It is estimated from sample data and is not available for all small areas. C is normally the census count, including imputations and erroneous inclusions (duplicates, etc.). Only the Census count is available for all areas.

So, using the synthetic model

$$\hat{N}^s_{jkh} = CCF_j C_{jkh} \qquad (4)$$

Summing over group and subclass yields a measured population for a given geographic area (state, county, town).

$$\hat{N}^s_k = \sum_j \sum_h CCF_j C_{jkh} \qquad (5)$$

For example, $j$ may define all 0-18 year old Asians in the West region, while $k$ may define Orange County, California and $h$ may define 11 year old girls.

While this produces a small-area and small-group estimator, this calculation can generate fractions. The typical public user of census data prefers whole person records. We use controlled rounding and person imputation to create integral number of person records for ease of tabulation and data acceptance.

## 3. A.C.E. Operations and Timing

The purpose of the A.C.E. design and operations is to provide the data needed for the model presented above. This section presents an overview of A.C.E. operations and timing. It also notes some of the principal changes from the 1990 PES. For a more detailed analysis, the reader is referred to Childers and Fenstermaker [2000], Childers [2000].

The A.C.E. operations begin by creating a universe of block clusters that cover the entire land area. This is done by combining contiguous census blocks together into block clusters containing approximately thirty housing units, as measured by early census address lists. Even blocks with no listed housing units are part of the universe. However, when possible small blocks are attached with large blocks into clusters.

A stratified random sample of these block clusters is selected. The E-sample is stratified by state, by size of block cluster and by 1990 demography. Cluster size was classified into small (two or fewer housing units), medium (3 to 79 housing units) and large (80 and more). Measures of size are based on early (1999) census housing unit counts. The clusters were also stratified by race and ethnic composition (Asian and Pacific Islander, Hispanic, Black and All Other) and proportion of owner occupied housing units based on 1990 Census data. A separate stratum was kept for American Indian reservations. The overall sample contained approximately two million housing units for listing operations.

From September to November 1999, A.C.E. address listers visited all sampled block clusters. They were given a "blank" map showing streets and cluster boundaries but no census addresses. They were asked to list all housing units and potential housing units, e.g., trailer pads and construction sites. This work was quality controlled. In 1990, PES address listing was conducted in January and February. The change allows for better weather during listing and also the inclusion of initial housing unit matching (see below). It also allows a more efficient sample design.

In January 2000, the listed sample block clusters were subsampled (sample reduction). For sample reduction, the clusters were restratified based on the A.C.E. housing unit count, the updated census housing unit count, the difference between these, together with the 1990 demographic composition again.

In an operation new since 1990, the housing units independently listed in the reduced sample clusters are matched to the census housing units for the clusters. Computer and computer-assisted clerical matching was used, with further field verification where called for. The purpose of the housing unit matching was to create an accurate linked list of the housing units in the block. This list is used in subsequent operations including large block subsampling, telephone interviewing and person matching. However, neither the results from the census nor those from the A.C.E. are carried over to "improve" the other list. Although linked, the lists are kept separate.

Approximately 300,000 housing units will be selected for A.C.E. interviewing. In contrast, the 1990 PES selected 165,000 housing units. The A.C.E. universe excludes the population living in institutions, college dormitories, and other group quarters. The 1990 PES included most non-institutional group quarters.

The A.C.E. interviewing will begin in May 2000. Interviewing will be done using laptop computers. (All 1990 interviewing was done with pencil and paper.) During the first few weeks, interviewing will be done by telephone. This phase includes only single-family housing units with house numbers and street name addresses that mailed back their census questionnaires and included a telephone number. This allows us, through the results of the housing unit matching phase, to verify that the unit has already been enumerated and to obtain the telephone number. This telephone interviewing is decentralized using the same instrument and staff as personal visit interviewing. Telephone interviewing was not conducted in 1990.

Because of the restrictions placed on the telephone interview universe, we expect that most A.C.E. interviewing will be conducted through personal visits from July through August. This time period is chosen because it follows the completion of census nonresponse follow-up. Census coverage improvement follow-up will overlap with A.C.E. interviewing, but affects less than one percent of the census housing unit universe. The 1990 PES was conducted during the same time period.

The A.C.E. whole household nonresponse cases will be returned to the field in September for another attempt called the nonresponse conversion operation. This operation was also conducted in 1990. However, the smaller sample size and other factors allowed the 1990 operation largely to use permanent Census Bureau Field Staff. This will not be the case in 2000.

A major change since 1990 is the treatment of those who move between Census Day (April 1) and the time of the A.C.E. interview. In 1990, the PES attempted to list all current household residents and ascertain their usual April 1 residence. In A.C.E. 2000, interviewing will focus on reconstructing the Census Day household. People who moved in will be listed, but the interviewer will not attempt to determine their location as of April 1.

Person matching begins using a computer matching program similar to that used in 1990. The results (matches, probably matches and nonmatches) are sent to clerks for resolution. Since the A.C.E. interview was computer based, the clerical matching is essentially paperless. It has also been made more user friendly through "windows" technology. Computerization of both the Census and the A.C.E. now allows all matching to be done in one site (in contrast to seven sites in 1990).

Because we are only matching people who resided in the sample cluster on April 1, the matching is greatly simplified compared to 1990. The entire mover matching operation needed in 1990 has been eliminated.

As in 1990, cases requiring additional information are visited after matching reinterview (follow-up). After follow-up, matching completes the process.

Nonresponse and unresolved cases are estimated through the missing data procedure. The details of this are not fully specified at this time, but its overall functional requirements will be similar to 1990.

As in 1990, the sample data will be weighted inversely to the sampling fraction and the DSE computed. The different treatment of movers requires a modification of the DSE as discussed below. In addition, the post-stratification variables (the *j*'s) have been somewhat modified. (See below.)

In 1990, the computed DSEs were smoothed using a composite regression estimator for the initial (July 1991) estimates. This approach was not used for later 1990 estimates, nor will it be used in 2000.

The estimated post-stratum undercount or overcount estimates will be distributed to the blocks proportional to the census total for that post-stratum and block. These numbers will be converted to integer values using a controlled rounding procedure similar to 1990. Individual person records will then be added to the census files using a hot-deck procedure. Individual person records representing measured undercounts and measured overcounts will be kept separately.

The following three sections show how these operations relate to the dual system model presented in Section 2.

## 4. Correct Enumerations

### 4.1 Definition

The first step in operationalizing Equation 2 is to define precisely what it means to be captured "In the Census." This list or set is defined to be those people "correctly" in the census. In this context "correctly" has four dimensions:

1. Appropriateness
2. Uniqueness
3. Completeness
4. Geographic correctness

"Appropriateness" simply means that the person should be included in the census. People who die before or who were born after the census reference date (April 1) are not part of the population (universe) to be measured. Similarly, records that refer to fictitious "people," tourists, or animals are out-of-scope.

"Uniqueness" refers to the fact that we wish to measure the number of people included in the census, not the number of census records. If more than one record refers to a single person, the count of records must be reduced for purposes of the DSE.

"Completeness" means that the record must be sufficient to identify a single person. If the record lacks sufficient identifying information, then we cannot determine either whether it was appropriately and uniquely included, nor whether it was also included in the second system, the PES.

Although completeness is necessary for the DSE, the census count includes imputations and other incomplete enumerations. Census operations normally have a requirement for a "data defined person." In Census 2000, the requirement is:

Two characteristics are required for a person to be data defined, where name counts as a characteristic. Name must have at least three characters in the first and last name together. The characteristics that are included in the counting are relationship, sex, race, Hispanic origin, and either age or year of birth.

When a data-captured record does not meet these requirements census processing substitutes (imputes) another data-defined record for these "non-data defined" persons. Since the census processing flags all these whole-person imputations, the quantities are known with certainty and need not be estimated. Traditionally, the number of whole person imputations is denoted by II.

Additionally, there are person records that are acceptable for census processing but insufficient for use in the DSE. This group includes, for example, records with reasonably complete data but without a person's name. Clearly, accurate matching or follow-up is not possible for these cases. For A.C.E. 2000, the definition for "sufficient information for matching" is:

> The minimum amount of data required for the data defined census people to have sufficient information for matching and follow-up is complete name and two characteristics.

"Geographic correctness" means that the person is included in the census where he or she should be. Enumerations outside that area are, by definition, counted in the census but not correctly included in the census.

This concept is under the control of the DSE designer. Conceivably, one could define the correct location as anywhere in the nation. However, the bigger the "correct location" is defined, the larger the area that must be searched during the matching process. As the size of the search area increases, the complexity increases and the chance of false matches grows. Normally a smaller area is defined. Two dimensions must be defined to operationalize a smaller area. One must define:

1. Correct location
2. The area of search around the correct location

The "correct location" defines where, under the DSE residence rules, the person should be included. These rules may differ from the rules used in the census. The only requirement is that the location be uniquely defined and consistently applied during PES processing.

In the 1990 PES and 2000 A.C.E. the Bureau of the Census adopted the following rule:

> The person is correctly included in the census if he or she is included at the location where the person believes, at the time of the survey interview, to have been his or her usual residence as of April 1.

Note that the definition generally follows the census rules. However, it makes an explicit allowance for the fact that where the person considers his/her usual (April 1) residence may have changed by the time of the survey interview. This, by itself, does not bias the DSE. However, it does, as we shall see, require consistent reporting of the "correct location."

The second dimension of geographic correctness is the area of search around the correct location, i.e., the search area. This is largely to accommodate errors in either the census or survey assignment of residents to a particular geography. It has the effect of lowering the variance and can, in some circumstances lower the bias as well.

In 1990, the search area was defined, in urban areas, as the block cluster containing the persons "correct location" plus all blocks that touched the cluster (the first ring). In rural areas it was expanded to a second ring. In remote (List/Enumerate) areas, the entire enumeration district was searched. Thus we could say, with respect to geographic correctness, that in the 1990 PES, an enumeration was correct if the person was counted in the block he or she (would have) reported as his usual residence or in any of the surrounding blocks.

For Census 2000, we have adopted the following definition: An enumeration will be correct if the person was counted in the block containing his/her usual residence, unless none of the census enumerations from that housing unit were in the block. In this case the enumeration would be correct if found in a surrounding block (is one ring) where the address is located.

The above concepts are used to define the number of people correctly included in the census $\left[N_{+1}\right]$.

Notice that this definition does not depend on the correctness of classification *j*. For example, if a person was really 19 years old, but was counted in the census as 17, he/she is still considered as correctly included. This is discussed in a later section.

## 4.2 Measurement

To actually estimate the number of people correctly included in the census, we must take a sample of all data-defined census enumerations. This sample is called the enumeration (or E) sample.[3]

The records in the E-sample will be checked for completeness. The determination of sufficient information will normally be based on the scanned image of the census form. If no image is available, e.g., an Internet response, the ASCII data will be accepted. Errors can occur if the matching clerk defines as insufficient (perhaps using only the ASCII) a record for which information exists, or, of course, vice versa. Records that do not meet the DSE matching requirement will be coded KE .

Records are then searched throughout the search area to see if the person was counted more than once within the sample block (uniqueness). Duplicate search is done using both computer and computer-assisted clerical matching. If more than one record is found, the extra records are coded as duplicates. Duplicates are erroneous enumerations and are not considered correctly captured in the census. They are, in effect, removed from consideration for DSE. Errors in identifying duplicates are few but can occur through clerical (human) carelessness or more often because the name and characteristics appear slightly different in each record. However, should an error in duplicate match occur, one or both records would be sent to the field for verification. Most often either the problem is resolved, or one record is coded as erroneous (fictitious) rather than erroneous (duplicate), causing almost no effect on the DSE.

3/ In some of the literature, the entire census is incorrectly referred to as the "E-sample." This usage confuses the universe with the sample.

Appropriateness and geographic location cannot be determined from the census enumeration alone, but require additional interviewing.[4] However, if interviewing locates a member of the household, or an acceptable respondent who can confirm the person's existence and also that the person had his/her usual residence there on April 1, the enumeration is accepted as correct.

If the respondent reports that the person did not live there on April 1, the enumeration is considered erroneous and coded (EE ). Often, this is because the person responded to the census but moved before April 1 or, more often, the person moved in after April 1, but was enumerated by the census nonresponse follow-up.

Alternatively, the interviewers and respondents may determine that the person never existed or at least was never associated with the block. These are considered erroneous and thus not part of $N_{+1}$. This is easy with comic book names or animal names. It can be difficult in some cases to prove that a "person" was not real, especially in a large block.

Errors can occur in several steps at this stage. An important source of error arises from the need to accept proxy responses to verify many enumerations. If the proxy reports a different "correct" residence than the person himself would, an enumeration could be miscoded. It is usually, but not always, possible to show that a person (or cartoon character) never lived at a given address. However, since he could have lived somewhere in the block, it can be difficult in some situations to code the record fictitious rather than as simply an E-sample nonresponse case. We require the interviewers to find at least three knowledgeable respondents before coding a record as fictitious.

A final source of error needs to be mentioned here: errors due to the missing data model. That there can be no "whole household" nonresponse in the E-sample as the universe is by definition all individual data-defined census records. Given the design, there are no missing data with respect to being unique (duplicate search) and defined (sufficient information for matching). However, to the extent appropriateness and correct residence cannot be determined in the field, they must be determined by the missing data model. Obviously, to the extent such cases are not "missing at random" within an estimation class, the expected value of the missing data model will not appropriately reflect the truth.

4/E-sample follow-up is required in the U.S. because experience has shown that the census contains a non-ignorable level of erroneous enumerations. Ideally, this would be controlled and excluded in the enumeration phase. Some countries assume that census erroneous enumerations can be ignored.

After missing data estimation and sample weighting, we can estimate the number of people correctly counted in the census as

$$N_{+1} = (C - II)\frac{CE}{N_e}$$

$C$ = Census total records, including imputed, duplicate, fictitious, etc. (the Census count),

$II$ = number of whole-person census imputations,

$CE$ = weighted estimate of appropriate, unique, complete and correct enumerations,

$N_e$ = weighted E-sample estimate of total, including duplicate, fictitious, etc.,

This completes the estimation of the number of people "in the Census."

## 5. Measuring the Proportion of People Correctly Enumerated

Having defined the number of correctly enumerated people, the next step in DSE is to estimate the ratio $N_{1+}/N_{11}$, the inverse of the census coverage rate.

Conceptually, estimating the ratio entails (1) taking a sample of people, (2) determining whether they should be enumerated in the census, and (3) determining whether they were, indeed, correctly enumerated using the definitions defined in the previous section. If an unbiased sample can be drawn of people who should have been enumerated and if we can determine whether they actually were correctly enumerated (included in the census), then the DSE will produce asymptomatically unbiased estimates.

To the extent that each step can be approximately correct, the results will approach a nearly, or at least usefully, unbiased estimate.

The first step in the process is, normally, to draw a random area sample. In both 1990 and 2000 the survey has been undertaken in a random sample of blocks. Interviewers then canvass the block and prepare a list of people who should have been (correctly) enumerated. This list constitutes the population or P-sample. The (weighted) sum of the people on this list, denoted $\hat{N}_p$, estimates $N_{1+}$. However, it is not the number which is of interest, but the ratio of $N_{11}$ to $N_{1+}$, which we approximate by the ratio of correct matches $\left(\hat{M}\right)$ to $\hat{N}_p$.

Operationally, the (correct) census records are searched to see if the P-sample people were enumerated. The (weighted) number who were matched $\left(\hat{M}\right)$ estimates $N_{11}$. The DSE model will work if we can maintain or, at least, approximate:

1. Operational independence
2. Consistent Reporting
3. Accurate matching
4. Homogeneity within post-strata $j$

## 5.1 Operational Independence

Operational independence is the easiest assumption to approximate, but still requires vigilance. In Census 2000, the A.C.E. sample is drawn and the housing units listed before the delivery of the census questionnaires. Although personal contact is minimal, some people may react differently to the census because of their inclusion in survey listing. Further, survey interviews take place after almost all census enumerations are completed, but some contamination could occur. Great care is taken to prevent the same field staff from working the same area in both Census and A.C.E. and to prevent sharing of information. Still, some people may react differently to the survey because they were enumerated, for example, by a very polite or very surly enumerator.

More insidiously, contamination can occur during the office processing. For example, in some surveys, clerks are allowed to first attempt to match cases and to then decide whether the survey interview was complete and sufficient. Clerks react to this by matching all the cases they can, even those with marginal information, and declaring as "nonresponse" all or most nonmatched cases. In other examples, all P-sample cases that can be matched are considered "correctly enumerated" while all nonmatched cases (and only nonmatched cases) are sent out for further field work (follow-up). Cases that cannot be successfully contacted a second time are excluded from the P-sample. Thus, a given case might be accepted as a complete, accurate, and sufficient P-sample interview if it matches to the census but would be excluded if it was not matched. Thus, the chances of being included in the survey is made operationally dependent on whether the case was included in the census, which violates independence.

Census 2000 guards against unnecessarily introducing operational dependence by adopting the philosophy that one first decides whether a case is acceptable for matching and only then attempts to match to the census. In other words, the philosophy is "Do not attempt to find a match unless you would be satisfied that, if no match is found, the person was not enumerated!"

Before beginning the matching, P-sample records are reviewed for:

1. Appropriateness
2. Uniqueness
3. Completeness
4. Geographic correctness

The P-sample interviews and census enumerations are screened for "sufficient information for matching" before matching. The clerk cannot match to a census person until they update the characteristics on the matching file using the image. It is impossible to match a P-sample person to a non-data-defined census person (II).

Further, the A.C. E. 2000 does not adopt the philosophy that only one interview is required to determine that a person was correctly enumerated but two interviews are required to determine that a person was not enumerated. Not all nonmatched cases are sent to follow-up. Specifically the emphasis throughout is not on getting the best estimate of $N_{1+}$ (e.g., the "best" A.C.E. coverage) but on the best estimate of the ratio $\left(N_{1+} / N_{11}\right)$.

The A.C.E. will contain almost no obviously fictitious records. We have instituted a quality assurance process to minimize other sloppy or dishonest A.C.E. interviewing. In addition, one important exception to our "no follow-up" rule are cases where A.C.E. fabrication is possible, e.g., cases where no one in the household matches implying possible fabrication. In addition, out of scope records, e.g., group quarters, are screened out. Occasionally, survey duplicates occur and these are eliminated (uniqueness). Finally, as alluded to above, if the survey interview does not meet minimal standards, the case is converted to nonresponse.

## 5.2 Consistent Reporting of Residence

To measure the number of people correctly included in both systems, we must determine whether or not a P-sample person was correctly enumerated. This is done by searching the correct census records for the area where the person should have been enumerated.

The first issue in matching is determining whether we are searching in the correct geographic location. The same definition of geographic correctness must apply both to whether an enumeration was correct (in the E-sample) and to whether the person was correctly enumerated (in the P-sample). Failure to make these concepts agree is termed "balancing error".

Specifically, we must have the same definition of "correct" location and the same search area around the correct location. Specifically, we adopt the rule referred to above:

> "The person is correctly included in the census if he/she is included at the location where the person believes, at the time of PES interview, to have been his/her usual residence as of April 1."

Errors in applying this rule can result in both erroneous non-matches and erroneous matches. Most obviously, consider a person enumerated at his/her April 1 usual residence, who moves and is interviewed by the A.C.E. in July at the new address. The DSE works if the same definition of location is used for both the P and E-samples.

For example, consider a  person who has two addresses but consistently reports one address as his usual residence (even if his response does not follow the letter of census rules). If he was interviewed in the E-sample at either address, he would report the same address as being his usual one. If he fell into the P-sample, he would report also the same address. Then, regardless of which address fell in sample, we would correctly and consistently classify him.

Difficulty comes primarily from two problems. First, both the P-and E-sample accept proxy responses. Thus, even though the person might have a clear and consistent understanding of his usual residence, the proxy respondent may not. The neighbor at each address may report that address as the person's correct address. Additionally, the way in which the question is posed in each interview could lead to different responses even from the same person.

If, say, a proxy misreports in the P-sample the current address as the person's usual April 1 residence, we will search the wrong area and, in this case, find a false nonmatch/not correctly enumerated. On the other hand, consider the case when the person was missed by the Census at his April 1 usual residence but incorrectly included by the census at the new address. Again, following the proxy, we would incorrectly count the person as "correctly enumerated." Although the person matches, he was not correctly enumerated. Proxy P-sample responses that do not match are sent to follow-up.

The important concept to remember at this point is that for the DSE to work, for a match to be correct, it must not just be to the correct person, it must also be at the correct location. A correct match must be to a correctly enumerated person.

The other dimension of geographic correctness is, again, the extent of search. Clearly, the area around the correct residence must be the same to determine both whether an enumeration was correct and whether a person was correctly enumerated. This is fairly easy to achieve by consistently applying the same definitions of the search area as in Section 3.

### 5.3 Accurate Matching

As mentioned above, the purpose of matching is to determine whether a person interviewed in the P-sample was also enumerated in the census within the defined search area. All indications are that the current matching system is very accurate in finding and determining a link if one exists.

Much of the matching is now done by a computerized matching system developed at the Census Bureau over the past sixteen years. The system outputs linked records (matches), possible matches and nonmatched cases. Repeated tests have shown that cases matched by the computer are quite likely to be correctly linked. (See for example Belin, 1993.) Meanwhile, nearly all clerical matching is now computer-assisted and largely paperless. This new system makes searching easier, including duplicate search. It restricts the codes clerks can apply to only those appropriate for the situation. Since the searching and data entry is now easier, we feel it is more likely to be done accurately. For example, the almost paperless system should eliminate lost and misfiled A.C.E. questionnaires.

The first-level clerks are backed up by a team of around 50 technicians. These technicians were hired this past summer and have been trained since September. They are backed up by a team of seven permanent analysts, most of whom have been matching for several years and, in some cases, decades.

Each level of matching acts as quality assurance for the lower level. In addition, each level can refer problem cases to the next higher level.

One big improvement in the matching has been made possible by the increased automation in the census. All matching will be done in one location by one staff. So even if biases should occur, they will not impact the results geographically. The 1980 and 1990 matching operations were done in three and seven sites, respectively.

The use of the A.C.E. procedures for movers (referred to as PES-C see below) also greatly simplifies the matching. Under the procedures used in 1980 and 1990 (referred to as PES-B), it was necessary first to locate and code the correct Census Day residence before beginning matching. This can be a difficult procedure, especially for example in rural areas. Mover matching was never automated.

In A.C.E. 2000 all matching, including for movers, will be done in the E-sample block, using the same computer and computer-assisted clerical matching system. For these reasons, we feel confident that our matching system will find and correctly link all matching records that are to be found, with a very low false match rate.

## 5.4 The Role of After-Matching Reinterview

Some cases are sent to the field to gather further information after the initial matching is complete. This after-matching reinterview is often termed the follow-up interview.

The follow-up interview process, like all PES activities, must fit into the overall framework of the DSE. Specifically, it must account for:

1. Requirement for appropriate, unique and correct response
2. Independence between census and survey inclusion probabilities
3. Balancing P-and E-sample errors
4. Unique location matching rules
5. Treatment of missing data.

Follow-up is only warranted if it provides better, i.e., more accurate or more consistent, responses. Simply obtaining a different response is not sufficient. Since follow-up takes place almost, by definition, further from the census reference date than the initial interview, it is more difficult to obtain accurate responses. This is equally true for E-sample follow-up and P-sample follow-up. In order to provide better responses, follow-up must use better resources, for example:

1. Better respondents (household vs. proxy)
2. A better trained, supervised or quality-controlled interviewer
3. Better questions or interview procedures.

For example, most people would agree that it would make little sense, all other things being equal, to replace an April or July interview of a household member with an October interview with a proxy. However, properly conducted, follow-up can obtain better information. The census data collection period extends from mid-March through mid-summer. Little emphasis is placed on verifying that the people were residents of the household on April 1. Because of this, we believe that a follow-up interview, with better trained and supervised interviewers and with probing questions, can do a better job than the census at obtaining what the people believe to be their usual residence as of April 1. In general, the A.C.E. has adopted the view that because of better training and supervision, and more complete questioning, the A.C.E. follow-up interviewing can, in general, obtain more accurate information on residence and location than that gathered during the census process itself. Thus all nonmatched E-sample cases are sent to follow-up.

Follow-up can, however, compromise independence. If all cases (or a sample of all cases) were sent to follow-up, independence would not necessarily be compromised. However, cases that are matched during initial matching are seldom sent to follow-up.

To do so would stress the resources available for follow-up. Instead, only non-matches or probably matched cases are usually selected for follow-up. This can introduce operational correlation bias.

Consider a situation where a woman lived in location A on April 1, moved into a sample area (location B) and incorrectly reported her usual census day address as location B. If she had been enumerated (incorrectly) by the census in location B, her records would match. The case would not be sent to follow-up, although it is, in fact, an incorrect match. However, if she was not enumerated in location B, she would not be matched, and her case sent to follow-up. Were she then to correctly report her Census Day location, she would be excluded from sample at location B. The net effect is that her inclusion in the survey is made to depend directly upon whether or not she was enumerated at location B. This constitutes operational dependence. Note that the bias occurs because the follow-up obtained the "correct" address.

The situation is somewhat mitigated if the incorrect enumeration to which she is matched falls into the E-sample. Then, if we allow the "false" match to stand, we have created both a correct enumeration (E-sample) and match (P-sample).

Following up the matches would convert the match to out-of-scope and the correct enumeration to an erroneous enumeration. So at least the number of correctly enumerated people and number of correct enumerations would have both been reduced by one. The balancing of the P-and E-sample errors mitigates but does not eliminate the bias of selective follow-up.

The above example assumed that the woman was incorrectly enumerated in the E-sample block. The same dependence can occur even for a correct enumeration. Assume that the woman has two homes, a vacation home in location A (where she spent April) and a usual home in location B (where she spends most of the year). In the survey she might well correctly report location B as her usual residence. If she was correctly enumerated at location B and correctly matched, she is not asked again. If she were missed by the census, she would be interviewed again. Trying to be helpful (after all, we didn't like her first answer) she might well now report location A, causing her to be removed from the survey. Of course, no bias is introduced if she were to report consistently during all interviews. However, in that case there is no point in reinterview.

As we have discussed above, the correct or unique location approach depends upon allowing matches only to correct enumerations. A follow-up interview (for non-matches) that produced an alternative new address might well produce a match at the alternative address. However, since matching was conducted on more than one location, we are violating this principal. This situation was more likely to occur in the 1980 and 1990

PES which interviewed current residents and sought through interview and reinterview to obtain this correct location. Because A.C.E. 2000 does not gather alternative addresses to follow-up this problem does not arise.

The biases that can be introduced by following up non-matches discussed above can occur even if the follow-up interview is successful. If the follow-up operation results in a non-interview, further biases can be introduced depending upon the missing data models applied to these cases.

When only nonmatched cases are sent to follow-up, obviously the follow-up universe is enriched with people who have not been correctly enumerated. The missing data model must account for this. Specifically, nonresponse follow-up cases that have previously been subjected to matching should only be imputed (or modeled) based on response follow-up cases that have also been subjected to matching. Follow-up nonresponse models that used initially matched cases to predict the enumeration status of cases that did not initially match can introduce a very powerful bias. This bias is similar mathematically to throwing out of sample a proportion of nonmatched cases because they did not match.

A careful nonresponse model can lessen the problems of follow-up noninterview. However, as in all missing data procedures, one is left, at some level, with the assumption that the observed cases are like the unobserved cases. This will be only approximately true.

One clear danger is that the very hardest-to-count people might, with some luck or effort, be enumerated or interviewed in the initial process. Going out months later to find them again might well result in a noninterview. Easier to count people would be easier to follow-up. Thus follow-up might well introduce correlation bias even after missing data modeling.

Choosing cases for follow-up then requires balancing the needs for accurate and consistent information with the need for independence and consistent definition of correct enumeration.

As noted above, for E-sample nonmatched cases the A.C.E. adopts the view that the better staff, training, and questions mean that the gains from improved reporting outweigh the losses in independence.

For the P-sample, our philosophy is to send out cases only where better information is likely. Cases sent to follow-up include:

1. Possible matches, since with the information at hand the interviews can resolve the situation.

2. Initial proxy interviews that result in nonmatches. Since we have not spoken to a household member, we have reason to doubt the accuracy.
3. Nonmatched cases where, for the same housing unit, the census reports one family and the A.C.E. reports another. In order to ensure consistent reporting of census day address between the P-sample and the E-sample, these cases are sent out together.
4. Partial household nonmatches.

Cases that match and other nonmatched cases are not sent to follow-up. Although the missing data procedures for A.C.E. have not been fully specified, it will model nonresponse cases that have been subjected to matching only from response cases from the same universe. (See Cantwell, 2000)

### 5.5 Homogeneity Within Post-strata *j*

As noted in Section 2, the usual Peterson or Chandrasekar-Deming DSE requires that the capture probabilities be independent for all individuals within estimation domains called post-strata. This is usually approximated by making the post-strata as homogenous as possible with respect to the census capture probabilities and then striving for as uniform as possible inclusion probabilities for the survey.

### 5.5.1 Poststratification

For A.C.E. 2000, we have conducted extensive research in defining the post-strata. (See Griffin and Haines [2000].) In addition, we have factored in what we know about changes between 1990 and 2000 and our experience in the 1998 Census Dress Rehearsal.

In defining post-strata for DSE estimation, we must balance the need for smaller more homogenous strata against an increase in sampling variance and in ratio bias. Ratio bias follows from the fact that the DSE is inherently a ratio estimator based on a sample of the population. In addition, our treatment of movers adds an additional ratio bias (see below). For this reason, we have designed post-strata with a minimum expected sample size of 100.

For A.C.E. 2000, we are currently focusing on post-strata based on the following variables:

1. Race and Origin (7)
2. Age and sex (7)
3. Tenure (2)
4. Urbanicity and type of enumeration area (4)
5. Mail Return rates (2)

The reasons we believe these to be important explanatory variables are as follows. All research and documentation of the undercount in the U. S. have demonstrated coverage differences between racial and ethnic groups. We believe that this is due to social, cultural, linguistic and economic differences between racial and ethnic groups that lead them to react differently to the census procedures. Thus, we believe that while people are not missed because of their race or ethnicity, race and ethnicity is correlated with other factors that lead people to be undercounted.

Demographic analysis and previous coverage surveys have demonstrated that people are differentially missed in different age groups and that the pattern is different for males and females. Most important in this pattern is that young adults are very mobile and often unmotivated by Census outreach and publicity.

The importance of tenure was first measured following the 1980 Census and was implemented in the 1990 post-stratification. We believe that those who live in owner-occupied houses are less mobile and thus easier to count. In addition, they may feel more of a stake in their community and thus are more influenced by the census outreach and publicity program.

Urban density is related to a number of variables. First, it obviously affects housing patterns and thus is correlated with address list development and housing unit coverage. Similarly, the way the Bureau of the Census builds its address lists is related to housing unit coverage. Thus, the combined variable " urbanicity and type of enumeration area" isolates differences in housing unit coverage. It may in addition, measure some aspects of social and economic isolation. Response to the census measures two important aspects related to coverage. First, it measures public cooperation with the census, an important predictor of coverage. Secondly, it measures directly the proportion of the enumeration that must be done in the census nonresponse follow-up. The larger the nonresponse follow-up workload, the harder it is to staff, train and supervise a temporary workforce. This can lead directly to coverage errors. One difficulty in this variable is that not all areas of the country use mail-back. A small proportion is done by direct interview. These obviously have no "mail response rate." Since, to an extent, the mail response rate measures unexpected and unusual difficulty, we have chosen not to group these areas with low mail response areas.

The variables for A.C.E. 2000 are similar to those used in the 1990 post-censal undercount estimates post-stratification. A major change we are considering is substituting mail-return rates for census region (North, Midwest, South and West). While the region is not entirely uncorrelated with undercount, we have become increasingly convinced that political geography, especially at such an aggregate level, is a poor predictor of either how census procedures are implemented or how people react to them. However our decision will be driven by the available data.

Obviously, the complete cross classifications can lead to very small cells. Appendix 1 gives the maximum set of post-strata we believe we can support. In planning these post-strata, we face the decision of how to classify people who choose more than one race category. Since this was not allowed in any previous census, we had very limited data upon which to decide. Appendix 2 gives our rules for the treatment of multiple race respondents together with the rationale behind these choices.

### 5.5.2 Treatment of Movers

As alluded to above, people who move present a special challenge for designing a DSE for census application. This flows from two considerations. First, people who move are more likely to be missed by the census and by the survey. Secondly, if a person has a different "usual residence" at the time of the survey than he did at the time of the census, one must decide where to sample him.

In the 1990 PES, movers were E-sampled where they lived at the time of the survey interview. We then searched the census records at, and only at, their April 1 usual residence. This is known as procedure (or PES) B. Although conceptually simple, the approach requires both coding the correct Census Day geography and then matching. These activities are complex and time consuming.

For Census 2000, a different procedure is used known as procedure (or PES) C. The A.C.E. 2000 will estimate the number of movers by the number of people who moved into the sample blocks between April 1 and the time of the A.C.E. interview (in-movers). If the population was closed to international migration, deaths, movement to group quarters, etc., then the number of people who moved in must equal the number who moved out (out-movers). The true total of out-movers should equal the total of in-movers. They are the same people in the population if not in the sample. It is normally easier to find people where they are.

The proportion of movers who are correctly enumerated is estimated by matching the out-movers to the census records for the E-sample block and extended search area, if appropriate. The estimated number of correctly enumerated movers is then

$$\hat{M}_t = \frac{\hat{M}_o}{\hat{N}_o} \hat{N}_i$$

where $\hat{M}$ denotes the weighted number of correct matches; $\hat{N}$ denotes the weighted population number, and the subscript denotes total moving ($t$) outmovers ($o$) and inmovers ($i$).

If we denote those who do not move by the subscript n, the overall coverage rate becomes

$$\frac{N_{11}}{N_{+1}} = \frac{\hat{M}_n + \hat{M}_t}{\hat{N}_n + \hat{N}_i}$$

The effect of procedure C is to increase the effective capture probabilities for movers and thus increase homogeneity with respect to mover status. (See Griffin 2000.)

Using post-stratification and reweighting movers can reduce the correlation due to heterogeneity in the post-strata. It will not be totally eliminated. This heterogeneity will cause the DSE to underestimate the true population.

### 5.5.3 Misclassification Error

In the discussion so far, we have accepted the post-stratum classification, "*j,*" as fixed. In practice, some people will be classified in different post-strata in the census and in the survey. For example, a woman may be reported as age 28 in the census and 31 in the survey, placing her in different post-strata.

Such misreporting is normally not important for matching. Name, address, relation and household composition are far more important than ages, race or sometimes even sex. So, assuming a match, in the above example we would have one correctly enumerated 28 year old in the E-sample and one correctly enumerated 31 year old in the P-sample. Misclassification can be seen to have two effects. To the extent the true undercount probabilities are homogeneous with respect to the true characteristics, misclassification introduces heterogeneity (and heterogeneity bias) into the observed estimation cells. Note that this is true even if reporting is consistent between the census and the survey.

Inconsistent reporting introduces another problem. Essentially, the coverage for group *j* as measured by the P-sample is misapplied to a different group in the DSE.

For example, see Table 1. In this example, there are two groups. One has 90 percent coverage by the census, the other only 80 percent. Thus, in this example, 1000 people from each group are included in the census. The survey correctly classifies all people and correctly measures their coverage by the census. We assume that 20 percent of the individuals from group 1 are incorrectly classified into group 2 by the census. Therefore, the 0.9 coverage ratio is applied to only 800 people. The 0.8 coverage ratio is applied to 1200 people.

As the reader can see, even in this extreme example, the estimated total population is only off by one percent. The population for the subgroups remain in error. However, correcting for classification error is not the purpose of the DSE.

It is sometimes suggested that a better (more consistent) classification can be had by recoding the matched case to agree. However, this could lead to an even more serious error since none of the non-matching cases would be recoded.

Table 1

| Panel A - Number in Census | | | | | |
|---|---|---|---|---|---|
| Group<br>($j$) | True Population | Coverage Ratio<br>($N_{11}/N_{1+}$) | True Classification | Census Classification | Observed DSE |
| 1 | 1111 | 0.9 | 1000 | 800 | 888 |
| 2 | 1250 | 0.8 | 1000 | 1200 | 1500 |
| Total | 2361 | | 2000 | 2000 | 2388 |

## 5.6 Missing Data

As in all surveys, there will be nonresponse and incomplete response at various steps. The goal of the missing data process is, obviously, to improve the estimate of the coverage ratio. In choosing missing data procedures, we choose methods that support the underlying DSE assumptions.

Missing data can occur with an initial failure to get a survey interview. One possible approach is to treat these cases as not in the survey, i.e., as in $N_{2+}$. This treatment is adequate if whole household nonresponse is not correlated with census coverage. To the extent it is, ignoring these cases increases the bias due to correlation.

If we believe that household nonresponse cases are similar to survey response cases within the same cluster, we can reduce the bias due to nonresponse by reweighting the cluster. In A.C.E. 2000, we will use two sets of nonresponse adjustments in the P-sample: one applied to non-movers and out-movers (for use in estimating the *proportion* of matches), the other applied to in-movers (for use in estimating the *number* of movers). For each procedure, if there are enough households, we will form nonresponse cells within the block cluster according to the type of basic address: single-family, apartment, and other. Where it is necessary, we will collapse cells across clusters of similar composition, or across the types of basic address--according to pre-specified rules.

Even if the household has been interviewed, some PES cases can be unusable, for example, those lacking sufficient information for matching and those whose residence status is still unresolved. When these are screened out before matching is attempted, one again has the option of treating them as "not in the survey," i.e., in $N_{2+}$. Again, one can do slightly better by taking advantage of what knowledge we have.

However, once matching has been attempted, it is no longer acceptable to treat resulting nonresponse cases as "not in the PES" $\left(N_{2+}\right)$,or otherwise to use a missing data mechanism that distributes the weights evenly among all resolved cases. To do so would be to seriously violate the independence assumption. Only cases that did not match would be subject to becoming nonresponse cases. However, they would at least implicitly be given the coverage rate of all cases.

An important class are cases (1) that do not initially match, (2) are then sent to the field for additional interviewing, but (3) for which the additional interview is not successful (failed follow-up). An efficient matching process will find most of the matches. Thus, cases sent to follow-up will be disproportionally, if not predominantly, weighted with cases that were truly not in the census. Most follow-up nonresponse cases are probably, in fact, not enumerated. Giving them the overall match rate is to assume that most are indeed enumerated.

In the A.C.E. 2000, for a P-sample person with unresolved residence status, we will assign a probability of being a resident on Census Day according to operational or demographic information collected on the person. The idea is to group together into imputation cells people who are similar with respect to the A.C.E. operations--matches needing follow-up, nonmatches needing follow-up from whole-household nonmatches, persons resolved before follow-up, etc.--or demographic characteristics. Within each imputation cell, we assign to the unresolved cases the weighted average residence probability of all resolved cases.

## 6.0 Synthetic Estimation

### 6.1 The Synthetic and Dual System Model

To this point, we have been dealing with the actual DSE. However, as noted in Section 2, we use a synthetic estimator to distribute the undercount to local areas and small groups as noted in Equations 4 and 5.

In A.C.E. 2000, the carrying down is based on the same post-stratification variables as the DSE itself. Obviously, one could combine DSE post-strata to produce the synthetic post-strata. The synthetic estimation is based on the assumptions (1) that the DSE estimates the true population, and (2) that (within post-strata) the true population is distributed proportionally to the preadjustment (expected) census count. We have discussed the underlying basis of the first assumption above, so here we will concentrate on the second.

Clearly, at some level the second assumption is only true with respect to the expected census counts. That is, even if within post-strata all people had identical probabilities of being enumerated in the census, we would observe different outcomes across blocks. The underlying DSE explicitly models the undercount as a stochastic process. That is for any area and any post-stratum, we may express the relation as stochastic.

$$C_{jkh} = \quad N_{jkh} \cdot \in_{jkh}$$

where $\in_{jkh}$ is a proportional error, following some distribution. The N denotes the true population. So (dropping the subscript)

$$\hat{N} = CCF \cdot C$$
$$\hat{N} = CCF \cdot N \cdot \in \quad = \quad N \cdot CCF \cdot \in$$

So $\hat{N}$ is an unbiased estimator of N if $E(CCF \cdot \in) = 1$

However, *CFF* is effectively a constant with respect to the coverage error in a single block. So for a given block, the level of correction will, obviously, not reflect the difference between the census and the truth, even when the estimator is unbiased.

However, this is only true for small areas. As areas get larger, two things happen. First, the stochastic effect, or the random "block effect" begins to average out. Secondly, the effect of the actual undercount from a collection of blocks becomes positively correlated with the post-stratum coverage factor. This is, the larger the area, the more the area's undercount determines the net correction factor.

A simple example can help to demonstrate this effect. Consider a post-stratum with three equally sized blocks (say 30 persons each). Assume that one block and all its residents were completely missed by the census. So the true population is 90, but the census only counted 60. Assume that the PES worked perfectly and estimated the true population at 90, calculating a coverage factor of 3/2. The synthetic estimator would work as follows:

Table 2: Synthetic Estimation Example

| Block | True Population | Census C | Coverage Estimator $CxCCF = \hat{N}$ | Census Error $C - N$ | Adjusted Error $\hat{N} - N$ |
|---|---|---|---|---|---|
| 1 | 30 | 30 | 30x3/2=45 | 0 | -15 |
| 2 | 30 | 30 | 30x3/2=45 | 0 | -15 |
| 3 | 30 | 0 | 0x3/2=0 | 30 | 30 |
| Total | 90 | 60 | 60 90 | 30 | 0 |

Looking at the individual blocks one must conclude that the adjustment made two worse without improving the count for any. However, summing the blocks to the entire area one must conclude that the estimate was an improvement.

The stochastic effect would be trivial for all but the smallest areas if Wolter's autonomous independence assumption held in practice. In fact, it is well known that within family or block people are often missed as a group. The whole building (or sometimes even block) might be missed by the census address listing procedure. Seldom do people refuse to respond as individuals. Rather, the household refuses as a unit. The failure of the antonomous independence assumption does not cause a bias in the dual system model as long as the underlying probabilities are equal within post-strata. However, when viewed post hoc for the purpose of the synthetic estimator, this failure can mean that observed coverage for a block is inconsistent with the estimated undercount adjustment.

## 6.2 Heterogeneity Error

As attention is turned to larger and larger areas the stochastic effect diminishes and is replaced with the problem of true heterogeneity of the underlying capture probabilities.

That is, even if the DSE were unbiased with respect to each post-stratum, there would be synthetic bias with respect to some local area.

To address the problem the A.C.E. 2000 has strived to select post-strata that approximate geographic homogeneity better than the 1990 PES. For example, within Hispanic post-strata, rural renters can be both recently arrived undercounted immigrants and military families living in rented base housing. By post-stratifying on census mail return rates, the A.C.E. design hopes to differentiate these two groups.

A related problem is how differential bias of the DSE by post-stratum might manifest itself through the synthetic estimator. For example, many people accept that, when matching is under control, the DSE will underestimate the population due to response correlation bias. When each post-stratum is viewed individually, one could say that the DSE moves the measure of population closer to the truth, just not all the way.

However, if response correlation bias is more pronounced in some post-strata than in others, then these post-strata would not move as close as they should. Members of these post-strata could conceivably be made relatively worse, although for many purposes, their measured population would be more accurate.

We might refer to the difference between the DSE (including the estimate of those missed by both systems) and the true population as the residual or unmeasured undercount. Obviously, we have little evidence of where these "unmeasured" people live. If correlation bias is so strong that the unmeasured people live in areas where the DSE measures the smallest undercount, the DSE/synthetic model may fail. However, if the unmeasured undercount exists in the same area as the measured undercount ( i.e., the two are correlated), we are likely to get an improvement.

## 6.3 Inconsistent Reporting

Inconsistent reporting between the census and the survey poses a problem for the synthetic estimator as well as for the DSE. This is easily seen by ignoring census imputations and erroneous enumeration. In this case, the coverage factor is the inverse of the matching rate $\left(N_{11j}/N_{1+j}\right)$. If the reporting of characteristic $j$ is inconsistent between the census and survey, we would be applying the rate, estimated from one group, to a somewhat different group. While misclassification may be ignorable at the post-stratum level, it may be important locally.

For example, assume that, as it is a face-to-face interview, the A.C.E. measures the undercount for American Indians who maintain tribal affiliation. However, it might happen that many more people who mail back their questionnaire self-identify as

American Indian. If the undercount is higher among the A.C.E. group than the census group, a significant error could be created. The impact of this error would no doubt play itself out differently in different local area.

The A.C.E. seeks to protect itself against the general problem by avoiding, when possible, post-stratum definitions based on variables with high reporting variability. For example, the A.C.E. mitigates the American Indian problem by creating a group of strata of American Indians living on reservations. American Indians included in this post-stratum are likely to have strong tribal and cultural affiliation.

## 7. Concluding Remarks

In this paper we have described the theory of the DSE, and have discussed how post-enumeration surveys in general, and A.C.E. 2000 in particular, have implemented that theory. We have described the approximations necessary in real applications and the types of errors that can occur.

Obviously, it is the role of the survey designer and survey manager to balance and minimize the errors so as to produce useful and accurate measures of the population. When this is successfully done, Census 2000 A.C.E. will produce fair and accurate population measures for use by American scholars, planners and leaders.

## 8.0 References

Belin, T. R., (1993), "Evaluating Sources of Variation in Record Linkage Through A Factorial Experiment." *Survey Methodology*, 19, 13-29.

Childers, D., (2000), "Accuracy and Coverage Evaluation: The Design Document" DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1

Childers, D., and Fenstermaker, D., (2000), "Accuracy and Coverage Evaluation: Overview of the Design" DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-2

Cantwell P., (2000), "Accuracy and Coverage Evaluation Survey: Missing Data Procedures" DSSD Census 2000 Procedures and Operations Memorandum Series Q-19

Griffin, R., (2000), "Accuracy and Coverage Evaluation Survey: Dual System Estimation" DSSD Census 2000 Procedures and Operations Memorandum Series Q-20

Griffin, R., and Haines, D., (2000), "Accuracy and Coverage Evaluation Survey: Post-stratification for Dual System Estimation." DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter Q-21

Haines, D., (2000), "Accuracy and Coverage Evaluation Survey: Synthetic Estimation" DSSD Accuracy and C overage Evaluation Survey Memorandum Series Q-22

Hogan, H., (1992), "The 1990 Post-Enumeration Survey: An Overview", *The American Statistician*, 46, 261-269.

Hogan, H., (1993), "The Post-Enumeration Survey: Operations and Results", *Journal of American Statistical Association,* Vol. 88, No. 423.

Marks, E. S., Seltzer, W., and Krotki, K. J. (1974), *Population Growth Estimation*, New York: Population Council.

Marks, E.S., (1979), "The Role of Dual System Estimation in Census Evaluation", in K. Krotki, Recent Developments in PGE, University of Alberta press, pp 156-188.

Peterson, C. G. J., (1896), "The Yearly Immigration of Young Plaice into the Limfjord from the German Sea," *Report of the Danish Biological Station,* 6, 1-48.

Sekar, C.C., and Deming, W.E., (1949), "On a Method of Estimating Birth and Death Rates and the Extent of Registration," *Journal of the American Statistical Association*, 44, 101-115.

U.S. Census Bureau, 1985, Evaluating Census of Population and Housing, Statistical Training Document, ISP-TR-5, Washington, D.C.

Wolter, K. M., (1986), "Some Coverage Error Models for Census Data," *Journal of the American Statistical* Association, 81, 338-346

**UNITED STATES DEPARTMENT OF COMMERCE**
**Bureau of the Census**
Washington, DC 20233-0001

January 11, 2000

DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-2

MEMORANDUM FOR Howard Hogan
Chief, Decennial Statistical Studies Division

From: Danny R. Childers DRC
Debbie Fenstermaker DF
Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation: Overview of the Design

Attached is a summary of the documentation of the design of the Census 2000 Accuracy and Coverage Evaluation. This is intended as an overview of the major steps of the operational and sampling design. For more details, see "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1. If there are any questions, contact Danny R. Childers (301-457-4184) or Debbie Fenstermaker (301-457-4195).

cc: DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
A. C. E. Implementation Team
Statistical Design Team Leaders
Chron

# The Design of the Census 2000 Accuracy and Coverage Evaluation

## 1.0 Introduction

The Census Bureau will conduct an Accuracy and Coverage Evaluation (A. C. E.) to measure the overall and differential coverage of the U.S. population in Census 2000. The A. C. E. will also provide base population figures for other Census Bureau programs, such as the Census Bureau's intercensal population estimates, American Community Survey, and other demographic surveys. Under the traditional census plan, the A. C. E. will not be used to adjust the census figures for reapportionment purposes.

The "Population Sample" or "P-sample" and "Enumeration Sample" or "E-sample" have traditionally defined the samples for dual system estimation. The P-sample consists of people enumerated independent of the census. The E-sample consists of people enumerated in the census. After matching and reconciliation, the P-sample yields an estimate of the population missed in the census while the E-sample yields an estimate of the correctly enumerated people in the census. Combining these two components yields an A. C. E. estimate of census coverage.

The major steps of the A. C. E. include various stages of sampling, matching, and field activities. See the Attachment for a flowchart of A. C. E. activities. Note that the sampling and operational activities are intermixed. In particular, most sampling activities occur before housing unit matching. After the initial A. C. E. sample is selected, the housing units within the sample block clusters are listed. The sample is reduced to provide the A. C. E. cluster design. Then, the A. C. E. housing units are matched to the census inventory of housing units. After reconciling the nonmatches, a list of A. C. E. housing units that are confirmed to have existed within the block clusters is prepared and large block subsampling is conducted. The housing units selected in sample define the P-sample of housing units. Person interviews are conducted for these P-sample housing units.

This memorandum outlines the major steps in the sampling and operational activities for A. C. E. For more details about the design of the A. C. E., see "Accuracy and Coverage Evaluation: The Design Document", DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1. Additional discussions of the missing data and dual system estimation activities can be found in "Accuracy and Coverage Evaluation Survey: Missing Data Procedures", DSSD Census 2000 Procedures and Operations Memorandum Series Q and in "Accuracy and Coverage Evaluation Survey - Dual System Estimation", DSSD Census 2000 Procedures and Operations Memorandum Series Q.

## 2.0 The Initial Housing Unit Phase

### 2.1 The Listing Sample

As a result of the January 1999 Supreme Court ruling against the use of sampling for apportionment, the Census Bureau redesigned the Integrated Coverage Measurement (ICM) Survey as an A. C. E. The ICM was planned as a 750,000 housing unit sample while the A. C. E. sample is planned to be approximately 300,000 housing units. By the time of the Supreme Court decision, earlier commitments had become operationalized based on the ICM sample design, and consequently, the A. C. E. sample design had to be derived from the ICM design. Therefore, the entire ICM sample was selected as originally planned and then reduced through various stages to yield the target housing unit sample size.

Under the ICM sampling plan, key features of the A. C. E. listing sample selection include:

- roughly equal sample sizes for most states except the most populous
- a separate sample for American Indian Reservations
- roughly proportional allocation of sample within states
- a separate sample of small block clusters
- an oversample of large block clusters

The A. C. E. primary sampling unit is the block cluster. A block cluster is a single census collection block or group of geographically contiguous census collection blocks. Using housing unit counts from an early census address list, block clusters were stratified by size: small (0 to 2 housing units), medium (3 to 79 housing units) and large (80 or more housing units). In states with a sufficient number of American Indians living on reservations based on the 1990 census count, a separate sampling stratum was formed of American Indian Reservation block clusters. Within each sampling stratum, a systematic sample of block clusters was selected with equal probability.

The initial listing sample was selected in the second quarter of 1999. This stage of sampling yielded 29,136 block clusters and roughly 2 million housing units to be listed in the 50 states and the District of Columbia. For Puerto Rico, there were 559 block clusters in the listing sample and roughly 50,000 housing units to be listed.

### 2.2 Independent Listing

An independent listing of the addresses of all the housing units in the A. C. E. sample clusters is conducted before census day. The A. C. E. housing units are the housing units recorded in the Independent Listing Books (ILB). Besides listing each housing unit in the cluster, the listers will inquire about housing units present at each special place and commercial structure to obtain any additional housing units.

After the listing books are received in the National Processing Center (NPC) in Jeffersonville, Indiana, they are checked-in and keyed. The quality assurance is 100 percent for the keying of these listing books. Keying rejects are reviewed clerically to correct errors before the matching begins. A data file of the A. C. E. housing units is created for matching. The A. C. E. maps will be scanned at the NPC for use in the housing unit matching.

### 2.3 Medium and Large Block Cluster Reduction

The A. C. E. cluster reduction implementation for medium and large block clusters is scheduled for December 1999 through mid-January 2000, following two major operations: the completion of the A. C. E. independent listing operation and an update of the census address list. The resulting national sample allocation will be roughly proportional to state population with some differential sampling within states.

One component of the A. C. E. cluster reduction design is stratifying the A. C. E. listing clusters based on the relationship of current housing unit counts from the A. C. E. independent listing and the updated census address list. Clusters will be differentially sampled in order to reduce the variance contribution due to inconsistencies between the census and the independent list. Clusters with significant differences between the counts are likely to have high erroneous enumerations and high nonmatch rates. The objective of differentially sampling these types of clusters is to reduce the magnitude of the weights associated with clusters having potentially high variance contributions.

Another component of the A. C. E. cluster reduction design is differentially sampling clusters based on the estimated demographic composition of the cluster. Clusters with a high proportion of a minority or Hispanic group are classified into a minority stratum. The objective of differentially sampling these types of clusters is to increase the sample size and improve the reliability of the A. C. E. population estimates for these subgroups.

The A. C. E. cluster reduction occurs before the start of housing unit matching in order to reduce the volume of clusters going into that operation. Following this reduction there are expected to be roughly 15,500 block clusters in the A. C. E. reduced sample for the 50 states and the District of Columbia. The number of block clusters in Puerto Rico is unchanged at 559 block clusters.

### 2.4 Small Block Cluster Reduction

The small block cluster reduction occurs in late January through early February 2000 following the completion of the keying of the independent listing books. Small clusters are expected to have between zero and two housing units based on the early census address list. Conducting interviews and follow-up operations in small block clusters is not cost-effective compared to large clusters. Therefore, to allocate A. C. E. resources more efficiently, a subsample of these small clusters will be selected for the A. C. E. interviewing sample.

The small block cluster reduction has three goals. The first goal is to avoid having small cluster weights that are extremely high compared to other cluster weights in the sample. The second goal is to have lower weights on small clusters where the number of housing units is different from the early census address list. These two goals attempt to reduce the contribution of small clusters to the variance of the dual system estimates. The third goal is to ensure that the field staff can efficiently manage the resulting workloads.

Using the keyed independent listing counts and the currently available census counts, the small block clusters within each state are stratified by size. Then, a systematic subsample is selected from each stratum with equal probability. All small block clusters which have 10 or more housing units, and are either List Enumerate or on an American Indian Reservation or other type of American Indian Country are kept in the A. C. E. sample.

Approximately 1500 of the 5000 small block clusters are expected to be retained in the A. C. E. sample. This is the last stage of cluster reduction; therefore, there is expected to be approximately 12,500 block clusters in the A. C. E. sample for the 50 states and the District of Columbia. For Puerto Rico, there will be roughly 500 block clusters.

### 2.5 Initial Housing Unit Matching

The housing units included on the census address list in January 2000 in the block clusters after the block cluster reductions are obtained for the housing unit matching. The objective of housing unit matching is to link the A. C. E. and census housing units for automated subsampling in large blocks for both the P-sample and the E-sample. In addition, the cleaned-up list of P-sample addresses makes the person interviewing easier because all housing units have been confirmed to exist as housing units. The addresses for housing units in the A. C. E. and the census housing units in the A. C. E. block clusters are first computer matched. The computer matching is followed by an automated clerical review. There is also a clerical search, which is limited to the block cluster, for duplicate housing units during this phase of the matching. Possible duplicates in both the A. C. E. and the census are identified.

### 2.6 Housing Unit Follow-up

The cases coded as not matched, possibly matched, or possible duplicates after the matching are sent for follow-up interviews for all types of basic street address codes. Selected matched cases are also sent for additional information. Specifically, the cases identified for field follow-up are:

- The A. C. E. addresses with a before follow-up code of not matched are sent to the field to confirm if they were housing units within the block cluster.
- The census addresses with a before follow-up code of not matched are sent to the field to confirm if they were housing units within the sample block cluster.
- The possible matches are sent to the field to determine if the A. C. E. and census addresses refer to the same housing unit.

- Census housing units that are identified as possible duplicates are followed up to determine if the two census addresses refer to the same housing unit.
- A. C. E. housing units that are identified as possible duplicates are followed up to determine if the two A. C. E. addresses refer to the same housing unit.
- Matched housing units with a unit status code of under construction, future construction, unfit for habitation, vacant trailer site in a mobile home park, and other to determine the status of the housing unit at the time of the follow-up interview, since changes may have occurred since listing.

The questions on the follow-up form are not designed to be read to respondents but are used as a guide for an interviewer. The answer to one question may be the result of asking several other questions. The interviewer appropriately modifies the questions, when necessary, to the situation that is encountered in the field and records the appropriate answers on the follow-up form. Many questions can be answered by observation.

After the field follow-up, the completed forms are returned to the processing office. Using the information obtained during the field work, an After Follow-up Match Code is assigned for cases sent to the field.

## 2.7 Reduction Within Large Block Clusters

When block clusters are large, it is necessary to reduce the housing units within the cluster to obtain manageable field workloads for A. C. E. interviewing and person follow-up without having a big impact on reliability. Further, it is important to geographically overlap the P- and E-samples to reduce the E-sample person follow-up workload.

### 2.7.1 P-Sample

Following the completion of the housing unit matching and follow-up operations, the reduction of housing units within large block clusters is done. Any block cluster with 80 or more independently listed housing units after the matching and follow-up operation is eligible for this reduction. After the reduction within large block clusters is done, the interview sample size for the fifty states and the District of Columbia will be roughly 300,000 housing units. This stage of subsampling reduces the number of housing units in a cluster to be in the A. C. E. interview sample.

The reduction of housing units within a large block cluster is done by forming segments of adjacent housing units and selecting one or more segments for A. C. E. person interviewing. The segments have approximately equal numbers of housing units within a block cluster. Segments of housing units are used as the sampling unit in order to obtain compact interviewing workloads and to facilitate overlapping P- and E-samples to reduce E-sample person follow-up workloads.

### 2.7.2 E-Sample

The source of the E-sample is the Census Unedited File (CUF) which will be available in the fall of 2000. All census records in the A. C. E. block clusters are eligible to be in the E-sample. However, to reduce the person follow-up workload, a reduction of the census housing units is done when there are more than 80 census housing units in the block cluster.

To create overlapping P- and E-samples within a cluster when the census housing units are reduced, the results of the P-sample reduction results are mapped onto the census records on the CUF in the block cluster. An overlapping P- and E-sample is not necessary but is desired for cost effectiveness. This is possible because when there was a correspondence between an A. C. E. independently listed address and a census address during the initial housing unit matching, the census identification number was assigned to the A. C. E. unit. If there are a significant number of census addresses which do not correspond to an A. C. E. unit, then a second systematic housing unit reduction will be done to reduce the E-sample person follow-up workloads.

## 3.0 Person Interview

Three types of people are collected in the person interview: those who lived at the sample address at the time of the interview and on census day (i.e., nonmovers), those who have moved into the sample address since census day (i.e., inmovers), and those who lived at the sample address on census day but lived elsewhere at the time of the A. C. E. interview (i.e., outmovers). Census day residence status is established within the Computer Assisted Personal Interview (CAPI) instrument for the nonmovers and outmovers. This information is needed to prepare the dual system estimates.

The A. C. E. person interview is conducted using a CAPI instrument. In order to get an early start for the interviewing, a telephone interview will be conducted for households where the census questionnaire is data captured and included a telephone number. Both households with mail returns and enumerator filled questionnaires are eligible for telephone interviews. Housing units without house number and street name addresses and housing units in small multi-unit structures will be excluded from the telephone interviewing. Large multi-units are included in the telephone interviewing because they tend to have unique unit designations. Many small multi-unit structures and rural areas do not have addresses that allow the telephone interviewer to distinctly identify the address. We do not want to interview any addresses on the telephone without unique addresses. All remaining interviews after the telephone operation is completed will be conducted in person. However, some nonresponse conversion operation interviews and interviews in gated communities or secured buildings may be conducted by telephone.

The person interview is conducted only with a household member for the first three weeks of interviewing. If an interview with a household member is not obtained after three weeks, an interview with a nonhousehold member is attempted, called a proxy interview. The proxy interviewing is allowed during the remainder of the interviewing period. During the last two weeks of interviewing a nonresponse conversion operation is attempted for the noninterviews using the interviewers considered to be the best available. This noninterview conversion will attempt to obtain an interview with a household member or a proxy respondent, but not a last resort interview[1]. This noninterview conversion will reduce the noninterview rate. .

After the names and characteristics are obtained, the resident status on census day is established. For nonmovers and outmovers, questions about mover status, group quarters, and other residences on census day establish the residence status.

## 4.0 The Person Phase

The P-sample people and the census people from the CUF are computer matched within cluster. The possible matches, P-sample nonmatches, and E-sample nonmatches are clerically reviewed using an automated computer match and review system. Additional matches and possible matches are identified by the clerical staff. Duplicates on both lists are also identified clerically. People with incomplete names are identified because they do not contain sufficient information for matching and follow-up. After the matching is completed, field follow-up is conducted for selected cases and the results of the field interview are coded in the matching database.

### 4.1 E-sample Insufficient Information for Matching and Follow-up

The census person records are reviewed to identify people with insufficient information for matching and follow-up. Only people with sufficient information for matching and follow-up are allowed to be processed in the matching and follow-up interviewing phases of the person matching. The three ways for a census record to be determined to have insufficient information are:

- The census people are not data defined.
- The census people are data defined, but computer coded as insufficient information for matching and follow-up.
- The census people are computer coded as sufficient information, but converted clerically to insufficient information for matching and follow-up.

The first type of census people who are not data defined are not included in the E-sample. Only data defined people are included in the E-sample. These data defined people create person records in the census. Sufficient information for matching and follow-up in the E-sample is complete name and two characteristics. See the design document referenced in Section 1.0 for a more detailed discussion of census data defined and insufficient information for matching and follow-up.

[1] A last resort interview is one obtained from a respondent who would not be classified as knowledgeable. This information may not include name and may contain minimal data.

### 4.2 P-sample Insufficient Information for Matching and Follow-up

The P-sample is reviewed by the computer software to identify people who have insufficient information for matching and follow-up. The P-sample rules for sufficient information for matching and follow-up are the same as the E-sample rules. Sufficient information for matching and follow-up in the P-sample is complete name and two characteristics. These cases identified by the computer will be suppressed from viewing by the clerical matchers to prevent errors in matching people who should not be converted to sufficient information for matching. The probability of matched will be imputed for the P-sample people coded as insufficient information for matching and follow-up. They are treated like other P-sample people with unresolved match status.

### 4.3 Person Matching

The people from A. C. E. housing units who will be initially matched to the E-sample and non E-sample census enumerations are:

- the nonmovers and outmovers identified as residents
- the people with unresolved residence status

The matching within the sample block clusters is done by the "computer matcher" followed by an automated clerical review. The computer matching will match the nonmovers and outmovers to the E-sample in subsampled blocks and then allow matching to the non E-sample census enumerations. These non E-sample enumerations are census people in housing units that were not included in the E-sample after the subsampling of census housing units. The goal of the matching is to produce the correct ratio of cases classified as omitted to those classified as included in the census.

Duplicates are identified in both the P-sample and E-sample people. A P-sample person duplicate is removed from the final P-sample. Whole households of P-sample duplicates are converted to noninterviews for Dual System Estimation. The A. C. E. interview was not a good interview when the whole household was duplicated. An E-sample person duplicate is an erroneous enumeration in the census.

### 4.4 Targeted Extended Search

The targeted extended search for 2000 A. C. E. is a two stage process. First, clusters are identified that will benefit most from expanding the search area to surrounding blocks because of geocoding error. Second, blocks within the cluster will be targeted for searching.

There are geocoding errors of exclusion and inclusion in the sample cluster. Geocoding errors of exclusion affect the P-sample nonmatch rate and geocoding errors of inclusion affect the E-sample erroneous enumeration rate. If the geocoding error omits the census housing unit from the sample block cluster, the P-sample people and housing units will not be matched. Conversely, if the geocoding error includes the census housing unit in the sample block cluster, the E-sample people will be erroneously enumerated.

The clusters selected for targeted extended search for the 2000 Accuracy and Coverage Evaluation are:

- Clusters included with certainty:
    - Relisted clusters in A. C. E.
    - Five percent of the clusters with the most census geocoding errors and A. C. E. address nonmatches
    - Five percent of the clusters with the most weighted census geocoding errors and A. C. E. address nonmatches

- Clusters selected at random from the clusters with A. C. E. housing unit nonmatches or census housing units identified as geocoding errors.

Clusters without A. C. E. housing unit nonmatches and census geocoding errors are out-of-scope for the targeted extended search sampling. The initial housing unit matching results are used to identify the A. C. E housing unit nonmatches and census housing unit geocoding errors. Any changes to the census inventory of housing units is not reflected in the housing unit matching used to identify targeted extended search clusters.

### 4.5 A. C. E. Person Follow-up

The person follow-up is conducted to gather additional information to accurately code the residence status of the nonmatched P-sample people and the enumeration status of the nonmatched E-sample people. The P-sample nonmatches do not match to the census. We want to make sure these P-sample nonmatches actually lived in the sample block cluster on census day. The P-sample nonmatch is sent for a follow-up interview when there is a possibility the residence status is not correct, such as partial household nonmatches, whole household nonmatches when the interview was obtained by a proxy interview, and when there is a conflicting household situation (i.e., Smith/Jones cases). The E-sample nonmatches are sent for a follow-up interview to determine if they were correctly or erroneously enumerated in the block cluster. We send possible matches for an interview to resolve their match status. There are also other cases sent to follow-up, such as matched people with unresolved residence status and other types of cases considered to have the potential for geographic errors in the P-sample.

The following table summarizes which P-sample nonmatches will be sent for a follow-up interview:

| Type of P-sample Nonmatch | Person Interview with a Proxy Respondent | Person Interview with a Household Member |
|---|---|---|
| Partial household nonmatch | Followed up | Followed up |
| Whole household nonmatch where the housing unit does not match | Followed up | Not followed up |
| Whole household nonmatch where the housing unit is matched to the census | Followed up | Not followed up |
| Whole household nonmatch with conflicting households | Followed up | Followed up |

In general, a partial household nonmatch is where there is at least one nonmatch and one matched P-sample person. A conflicting household is one where the address matches and both are occupied, but with different household members. For more information, see the detailed design document referenced in Section 1.0.

Also, we will add the following types of cases to follow-up to collect more information to verify P-sample housing unit geography in addition to asking the person nonmatch questions:

- All P-sample whole household nonmatches in relisted clusters, since they not included in the housing unit matching phase of the A. C. E.
- All P-sample whole household nonmatches in clusters with a high rate of P-sample person nonmatch. High has been defined as 45 percent.
- P-sample whole household nonmatches where the interviewer for the person interview changed the address for the P-sample housing unit. Information about the accuracy of the P-sample geography is obtained.
- Any P-sample whole household nonmatch identified by the Analyst as needing follow-up.

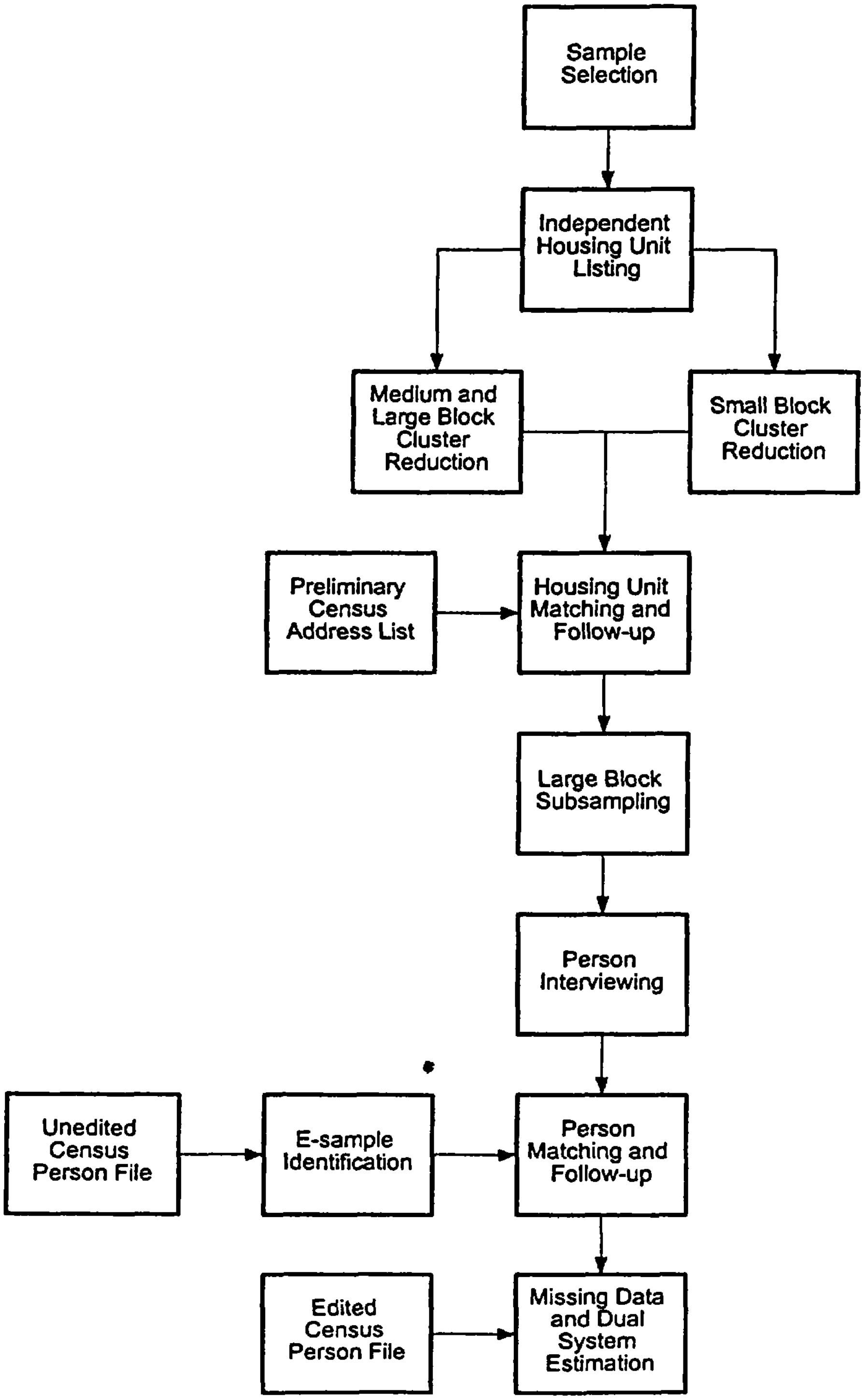
If the follow-up interview identifies a P-sample housing unit as a geocoding error, the people and housing units will be removed from the P-sample. The whole households of P-sample people incorrectly listed in the cluster will be coded as P-sample geocoding error.

The person follow-up is conducted using a paper questionnaire. The questionnaire is designed to gather information that may resolve matching and residence status problems. For example, a match between P-sample and E-sample people might be made if another piece of information is known. Also, information may be needed to confirm residence status on census day for matched or unmatched people. During the follow-up interview, interviewers will attempt to gather the information needed to code each person as a matched resident/non-resident or a nonmatched resident/non-resident of the block cluster on census day. The questionnaire places more emphasis on obtaining a good respondent before the follow-up questions are asked.

After the follow-up is completed, the results of the interview are reviewed and codes are entered into the system by the matching clerks. An outlier review is also conducted for clusters with high weighted nonmatch and erroneous enumeration rates. "Journals" will be written for these outlier clusters. There will be documentation for all outlier clusters.

The next step in the process is for the missing data work and the dual system estimation within post-strata. See the references in Section 1.0 for more details.

# Attachment: Workflow for the Accuracy and Coverage Evaluation


Sample Selection
Independent Housing Unit Listing
Medium and Large Block Cluster Reduction
Small Block Cluster Reduction
Preliminary Census Address List
Housing Unit Matching and Follow-up
Large Block Subsampling
Person Interviewing
Unedited Census Person File
E-sample Identification
Person Matching and Follow-up
Edited Census Person File
Missing Data and Dual System Estimation

UNITED STATES DEPARTMENT OF COMMERC
**Bureau of the Census**
Washington, DC 20233-0001

January 11, 2000

DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1

Memorandum For: Magdalena Ramos
Leader, A. C. E. Implementation Team

From: Danny R. Childers DRC
Leader, Design Team

Subject: Accuracy and Coverage Evaluation: The Design Document

Attached is the documentation of the design of the Census 2000 Accuracy and Coverage Evaluation. If there are any questions, contact Danny R. Childers (301-457-4184).

cc. DSSD Census 2000 Procedures and Operations Memorandum Series Distribution List
A. C. E. Implementation Team
Statistical Design Team Leaders
Chron

# The Design of the Census 2000
# Accuracy and Coverage Evaluation (A. C. E.)

Page

# The Design of the Census 2000 Accuracy and Coverage Evaluation

## 1.0 Introduction

The Census Bureau will conduct an Accuracy and Coverage Evaluation (A. C. E.) to measure the overall and differential coverage of the U.S. population in Census 2000. The A. C. E. will also provide base population figures for other Census Bureau programs, such as the Census Bureau's intercensal population estimates, American Community Survey, and other demographic surveys. Under the traditional census plan, the A. C. E. will not be used to adjust the census figures for reapportionment purposes.

The major steps of the A. C. E. are housing unit matching and person matching. Housing units within the sample block clusters are listed and matched to the census inventory of housing units. After reconciling the nonmatches, a list of A. C. E. housing units that are confirmed to have existed within the block clusters is prepared and person interviews are conducted for these P-sample housing units. Paper questionnaires are used for the housing unit listing and follow-up. The person interviewing is conducted by telephone and personal visit using Computer Assisted Personal Interview (CAPI).

Three types of people are collected in the person interview: those who lived at the sample address at the time of the interview and on census day (i.e., nonmovers), those who have moved into the sample address, since census day (i.e., inmovers), and those who lived at the sample address on census day but lived elsewhere at the time of the A. C. E. interview (i.e., outmovers). Census day residence status is established within the CAPI instrument for the nonmovers and outmovers. Questions are not asked about other residences to establish residence status at the time of the person interview for the inmovers. We obtain a proxy interview for the outmovers, since they have moved. No outmover tracing is conducted for the 2000 A. C. E.

The "Population Sample" or "P-sample" and "Enumeration Sample" or "E-sample" have traditionally defined the samples for dual system estimation. The P-sample consists of people enumerated independent of the census. The E-sample consists of people enumerated in the census. After matching and reconciliation, the P-sample yields an estimate of the population missed in the census while the E-sample yields an estimate of the correctly enumerated people in the census. Putting these two components together yields an A. C. E. estimate of census coverage.

There are non E-sample census people in large block clusters after subsampling. People counted in the census in institutional and noninstitutional group quarters are part of the non E-sample census people. Matching between the P-sample and non E-sample people is permitted, but non E-sample people who do not match are not sent for a follow-up interview. The correct enumeration rate is estimated from the E-sample. Matching between the P-sample and census people counted in group quarters is allowed in case there is misclassification of group quarters. If a P-sample person in a housing unit is not found in the census housing unit enumerations; the census people enumerated in group quarters are searched in case the person was enumerated in group quarters. Therefore, any misclassification of group quarters in either the P-sample or the census will not overestimate the nonmatch rate in the P-sample. After the computer and clerical matching, all E-sample nonmatches, selected P-sample nonmatches, and possible matches will be followed up in the field. In addition, people with unresolved residence status, even when they are matched, will be followed up to establish their residence status on census day. The person follow-up interview is a paper instrument.

Dual System Estimates for people in housing units are prepared. This approach, which has been called Procedure C, uses the best features of coverage measurement methodologies used in the past. Matching outmovers results in a better estimate of the match rate for movers than matching inmovers. The number of movers and their characteristics are better estimated from the inmovers than from the outmovers.

## 2.0 The Housing Unit Phase

### 2.1 Independent Listing

An independent listing of the addresses of all the housing units[1] in the A. C. E. sample clusters is conducted before census day. This list of housing units recorded in the Independent Listing Books (ILB) is defined to be the A. C. E. housing units. Besides listing each housing unit in the cluster, the listers will inquire about housing units present at each special place and commercial structure to obtain any additional housing units.

This listing is by basic street address initially. Each basic street address is assigned a map spot number and the map spot number is recorded on the A. C. E. map to identify the location of the basic street address. The address and other questions are asked by basic street address. The number of housing units within each basic street address is collected for each basic street address to increase the coverage of housing units in the A. C. E. This information is obtained by basic street address from the household member, by proxy, from apartment manager, or by observation. The basic street address is recorded once and the multi-unit coverage questions are recorded once. The individual housing units within basic street address are listed on the multi-unit pages of the listing book. Also, the A. C. E. lister is recording the number of units within basic street address on the map in parentheses to conform with census methodology.

Mobile homes that are not in mobile home parks are listed like single units. Each mobile home is assigned a unique map spot number and each mobile home is listed on a separate line in the listing book. If the mobile homes are in a park, the park is listed in the housing unit section of the listing book and each individual mobile home and vacant site is listed in the mobile home

[1] See Section 6.8 for the housing unit definition for Census 2000.

park section of the listing book.

Each individual mobile home is assigned a unique map spot number whether the mobile home is in a park or not. The location of the mobile home is identified by placing the map spot number for the mobile home on the map. This is the same procedure that is used by the census.

The following items are collected and recorded in the listing book for each basic street address:

- City style addresses (house number and street name)
- Non-city style address (route numbers, route and box numbers, or any other type of address that is not a city style address)
- Householder name (rural areas only[2])
- Description of address (for non-house number addresses only in both urban and rural areas)
- Number of housing units in basic street address
- Type of basic street address (single unit, multi-unit, mobile home not in a mobile home park, mobile home in a mobile home park, housing unit in special place, multi-unit in a special place, or other)
- Unit status for single units (intended for occupancy, under construction, future construction, unfit for habitation, boarded up, storage of household goods, and other)

The following items are also collected and recorded in the listing book for each multi-unit within basic street address:

- unit description/apartment number
- Unit status for multi units (intended for occupancy, under construction, future construction, unfit for habitation, boarded up, storage of household goods, and other)

The following items are also collected and recorded in the listing book for each mobile home in a mobile home park:

- house number, lot number, or physical description
- street name
- rural address
- Unit status (intended for occupancy, unfit for habitation, boarded up, storage of household goods, vacant trailer site in a mobile home park, and other)

After the listing books are received in the National Processing Center (NPC) in Jeffersonville,

[2] We need the household name for matching when the A. C. E. address is not house number and street name. We ask for the respondent name in all rural areas because the census address may not be house number and street name. We can not be guaranteed that the census address is collected in the same manner as the A. C. E.

they are checked in and keyed. The quality assurance is 100 percent for the keying of these listing books. Keying rejects are reviewed clerically to correct errors before the matching begins. A data file of the A. C. E. housing units is created to do the matching. The A. C. E. maps will be scanned in NPC to be used in the housing unit matching.

## 2.2 Sampling

As a result of the January 1999 Supreme Court ruling against the use of sampling for apportionment, the Census Bureau had to redesign the Integrated Coverage Measurement Survey (ICM) as an Accuracy and Coverage Evaluation Survey (A. C. E.). The ICM was planned as a 750,000 housing unit sample while the A. C. E. sample is planned to be approximately 300,000 housing units. By the time of the Supreme Court decision, earlier commitments had become operationalized based on the ICM sample design, and consequently, the A. C. E. sample design had to be derived from the ICM design. Therefore, the entire ICM sample was selected as originally planned and then subsampled through various stages to yield the target housing unit sample size.

The A. C. E. sample design inherited many of the ICM features: an American Indian Reservation sample, a small block cluster stratum and associated small block cluster subsampling, an oversample of large block clusters and complementary within large block cluster housing unit subsampling, and a separate sample for Puerto Rico. One key difference between the two sampling plans is the allocation of the sample. The ICM was geared toward direct state estimation thus requiring the 750,000 housing unit sample and roughly equal state sample sizes except for the most populous states. Further, the ICM plan was to allocate the sample almost proportionally within the states. The A. C. E. was not limited to a sample allocation to yield direct state estimates, but rather national coverage estimates allowing for the borrowing of strength across states. The A. C. E. sample of roughly 300,000 housing units is allocated to the 50 states and the District of Columbia proportional to estimated 1998 state population counts with a minimum sample size in small states and Hawaii. Within each state, the sample is differentially allocated among strata. This allocation is implemented through the A. C. E. cluster reduction operation.

The A. C. E. operations and the A. C. E. sample design are interrelated. First, because of the timing implications and infrastructure deployment that was underway for the independent listing operation, the A. C. E. sample design is contingent on the ICM sample design. A separate and independent 300,000 housing unit A. C. E. sample design was not feasible under the circumstances. Listing the entire ICM 750,000 housing unit design presented an opportunity to use double sampling techniques by making use of more current housing unit counts in stratification and updated measures of size. Therefore, the A. C. E. cluster reduction occurs following the A. C. E. independent listing operation, concurrently with the keying of the independent listing books, but before the housing unit matching operation begins. Further, the small block cluster operation must be done before the housing unit matching and follow-up since these clusters tend to require more per unit field resources than other clusters. The Housing unit matching plays a significant role in achieving overlapping P and E-samples, particularly in block clusters where the housing units are subsampled.

### 2.2.1 The Listing Sample

The listing sample was selected in the second quarter of 1999. It was important to select the listing sample by June, 1999, in order to produce the materials required to start the listing operation by September, 1999. Under this schedule it was necessary to use early census address list information to select the A. C. E. listing sample. The housing unit counts did not reflect all of the enhancements to the census address list which were incorporated by July, 1999 and did not include the final type of enumeration area classifications.

Under the ICM sampling plan, key features of the A. C. E. listing sample selection include:

- roughly equal sample sizes for most states except the most populous
- a separate sample for American Indian Reservations
- roughly proportional allocation of sample within states
- a separate sample of small block clusters
- an oversample of large block clusters

The A. C. E. primary sampling unit is the block cluster. A block cluster is a single census collection block or group of geographically contiguous census collection blocks. All census collection blocks in the 50 states, the District of Columbia and Puerto Rico are clustered except for collection blocks in remote Alaska areas. Remote Alaska is not in the sampling universe due to difficulties of the remote areas. The goal is to form block clusters which have identifiable boundaries and are of a manageable size in terms of both land area and number of housing units, roughly 30 housing units as indicated from the early census address list. Census blocks with 80 or more housing units were seldom clustered with other blocks except under specified situations. For A. C. E. 2000, small census blocks were allowed to be clustered with a neighboring census block containing housing units. This clustering was limited in order to prevent the geographical size of the clusters from becoming too unwieldy. Also, an important feature of block clustering is to create clusters which have similar census address list creation methods, and thus similar census types of enumeration areas. This is important operationally so that blocks within a cluster have similar census materials and identification methods (e.g., map spots).

Using the early census address list housing unit counts, block clusters were stratified by size: small (0 to 2 housing units), medium (3 to 79 housing units) and large (80 or more housing units). In states with a sufficient number of American Indians living on reservations based on the 1990 census count, a separate sampling stratum was formed of American Indian Reservation block clusters. Within each sampling stratum, a systematic sample of block clusters was selected with equal probability. Note that there was a workload limitation within each state on the number of housing units that could be listed. If the expected number of housing units to list exceeded the workload limit, then the number of sample clusters from the large stratum was reduced to maintain the listing workload targets.

This stage of sampling yielded 29,136 block clusters and roughly 2 million housing units to be listed in the 50 states and the District of Columbia. For Puerto Rico, there were 559 block clusters in the listing sample and roughly 50,000 housing units to be listed.

### 2.2.2 Medium and Large Block Cluster Reduction

The A. C. E. cluster reduction is planned to be implemented from December 1999 through mid-January 2000, following two major operations: the completion of the A. C. E. independent listing operation and the January update of the Decennial Master Address File (DMAF). *(If the DMAF is updated later, we will need to change this.)* The completion of these two operations provided two independent, up-to-date measures of size for the A. C. E. listing clusters. The preliminary independent listing counts were obtained following the quality assurance of the listing by clerically tallying directly from the independent listing books. These tallies were entered into the automated control system making these numbers readily available for use in the reduction. The updated DMAF housing unit count reflected the September and November 1999 Delivery Sequence File updates as well as the majority of the LUCA 98 field verification results.

One component of the A. C. E. cluster reduction design is stratifying the A. C. E. listing clusters based on the relationship of current housing unit counts from the A. C. E. independent listing and the census address list. Clusters will be differentially sampled in order to reduce the variance contribution of clusters that have inconsistent census and the independent list housing unit counts. Clusters with significant differences between the counts are likely to have high erroneous enumerations and high nonmatch rates. The objective of differentially sampling these types of clusters is to reduce the magnitude of the weights associated with clusters having potentially high coverage measurement implications.

Another component of the A. C. E. cluster reduction design is stratifying clusters based on the estimated demographic composition of the cluster. Clusters with a specified proportion of a racial or Hispanic group are classified into a minority stratum. The objective of differentially sampling these types of clusters is to increase the sample size and improve reliability of the A. C. E. population estimates for these subgroups.

A total of five reduction strata were formed within each state and the District of Columbia. The clusters were differentially subsampled within each state; retaining clusters from the minority and inconsistent strata at higher rates than the nonminority, consistent strata. Clusters in Puerto Rico, the small block cluster sampling stratum, and the American Indian Reservation stratum were not eligible to be subsampled.

The A. C. E. cluster reduction occurred concurrently with the keying of the independent listing books before the start of housing unit matching in order to reduce the volume of clusters going into that operation. Following this reduction there are expected to be roughly 15,500 block clusters in the A. C. E. reduced sample for the 50 states and the District of Columbia. The number of block clusters in Puerto Rico is unchanged, 559 block clusters.

### 2.2.3 Small Block Cluster Reduction

The original listing sample of 29,136 block clusters contained 5,000 clusters in the small block sampling stratum. Small blocks are over sampled. The small block subsampling selects blocks that are not small with certainty and a sample of 1 in 10 of the cluster identified as small to be included in the A. C. E. After sample reduction and small block subsampling there should be approximately 12,000 block clusters in the fifty states and the District of Columbia containing approximately 310,000 housing units. Small clusters on American Indian Reservations, in American Indian Country, and in list/enumerate areas are not being subsampled.

For small clusters not on American Indian Reservations or in American Indian Country, there will be three or four different subsampling strata: a one in ten subsampling rate for small clusters that remain small, a strata where all clusters with either 6 or more housing units or 10 or more housing units are retained, and one or two strata with intermediate subsampling rates. The listing and the DMAF counts will be used in defining the subsampling strata, instead of just the listing counts. There can be different small cluster subsampling rates in different states. The current expectation is that there will be a single set of subsampling rates for most states but a few states may have different rates to keep weights from getting too high.

## 2.3 Obtain DMAF Extract

The census housing units included on the Decennial Master Address File (DMAF) in January 2000 in the block clusters after sample reduction and small block subsampling are obtained for the housing unit matching. Changes made to the housing unit inventory in the DMAF after housing unit matching are processed during the final housing unit match.

## 2.4 Before Follow-up Matching

The addresses for housing units in the A. C. E. and the census housing units in the block clusters after sample reduction and small block subsampling are computer matched. The results of the computer matching are reviewed clerically. The clusters in Puerto Rico are not computer matched, but are matched clerically. The clusters in list/enumerate areas are not processed in housing unit matching. Processing for clusters in list/enumerate areas begins in Section 2.9 with large block subsampling. All housing unit matching in list/enumerate clusters is conducted in the final housing unit matching operation.

### 2.4.1 Computer Match

The housing unit data from the independent listing book (ILB) file and the DMAF extract go through a series of data preparation steps, including address standardization. Addresses from either file that are blank or could not be standardized are not computer matched, but these addresses are matched clerically.

The computer matching identifies the following match codes[3]:

M = The A. C. E. and census addresses match.
P = The A. C. E. and census addresses are possible matches.
NI = The A. C. E. address is not matched to a census address.
NE = The census address is not matched to an A. C. E. address.

A. C. E. addresses identified as mobile homes with no unit designation will be given a unit designation of "TRLR" for matching purposes. This prevents a mobile home at an address with no unit designation from matching to a census address representing a housing unit. Computer matching errors were discovered in the 1995 and 1996 tests when these type identical addresses were computer matched. The error is because the identification of mobile homes was not used in the computer matching. Adding TRLR keeps these computer matching errors from occurring.

The following criteria are used for improving the efficiency of the computer matching in multi-unit structures.

- Match all multi-units with alphabetic designations such as A, B, C, D, etc. to multi-units with numeric designations such as 1, 2, 3, 4, etc. In other words: A=1, B=2, C=3, D=4, E=5, ........... Y=25, Z=26.

- Match all multi-units with the following set of equivalent expressions:
  1) Apt 1=Apt A=Downstairs=Right=Lower=Front
  2) Apt 2=Apt B=Upstairs=Left=Upper=Rear

The results of the computer matching and additional information are loaded into a clerical matching support database which is utilized by the clerical matchers. This database is called the Matching Review and Coding System (MaRCS). The addresses will be in order by A. C. E. map spot number. In multi-unit structures, the individual apartments will be in within map spot ID order. The census housing units that do not match to the A. C. E. are in census ID order within block after the last A. C. E. housing unit.

[3] The A. C. E. addresses are not matched to census group quarters in the 2000 Accuracy and Coverage Evaluation. In dress rehearsal we matched to census group quarters. We may have caused damage to the P-sample in dress rehearsal. Also, there were many changes to the group quarters inventory after the special place enumeration. The P-sample people will be allowed to be matched to people enumerated in group quarters by including them with the other non E-sample people in the sample blocks in the 2000 Accuracy and Coverage Evaluation.

### 2.4.2 Clerical Match

The clerical matchers use the results of the computer matching to aid in their matching of addresses from the A. C. E. and the census. Only clusters expected to benefit from clerical matching are sent for clerical matching.[4] Supplemental materials are provided to facilitate the matching, such as the map spotted A. C. E. and census maps in the rural site. Before follow-up clerical match codes are assigned for each A. C. E. and census address that could not be matched by the computer. The matched addresses are not targeted for review, because the quality of the matches assigned by the computer has been proven to be good. Clerks are allowed to correct any errors in the computer matching that they identify while they are attempting to match the not matched housing units.

The clerical matchers use all housing unit information available to match housing units. The urban areas are almost totally city style addresses. Any addresses for housing units that are not matched for both the A. C. E. and the census after the clerical review are sent to the field for a follow-up interview.

In rural areas, the addresses are more difficult to match, predominantly because of the non-city style addresses. The city style address and the non-city style address are in two different locations in the census advance listing address register and on the DMAF. The matchers have householder names and location descriptions to help in matching for A. C. E. and census addresses in rural areas. The map spotted maps for the A. C. E. and the census in rural areas are used for the final determination that the housing units matched. Computer images of the A. C. E. and census map spotted maps which will be used in the housing unit matching are accessed by the Map Retrieval System (MRS) in MaRCS.

There is also a clerical search, which is limited to the block cluster, for duplicate housing units during this phase of the matching. Possible duplicates in both the A. C. E. and the census are identified. A follow-up interview is necessary to determine if the two addresses do in fact refer to the same housing unit.

The search area is limited to the sample block cluster in housing unit matching. A targeted extended search will be conducted for people in selected clusters during person matching. The search area will be extended for housing unit matching during the final housing unit matching on the clusters selected in person matching for extended search.

The goal for the 2000 A. C. E. is not to use any paper in the clerical matching. All materials needed for clerical matching should be available on the computer. Paperless matching reduces the time needed for clerical matching because the time spent waiting for an assignment and associated materials is eliminated. There is no need for a large staff to maintain an A. C. E.

[4] See section 2.5 for discussion on clusters sent directly to field follow-up without clerical review.

library with paperless matching.

The Before Follow-up Clerical Match Codes are:

M = The A. C. E. and census addresses are matched.

P = The A. C. E. and census addresses are possible matches. There is not enough information to assign a match with confidence.

NI = The A. C. E. address is not matched to a census address.

NE = The census address is not matched to an A. C. E. address.

DI = The A. C. E. address is a possible duplicate with another A. C. E. address. A follow-up interview is required to determine if the A. C. E. address is actually a duplicate of another A. C. E. address.

DE = The census address is a possible duplicate with another census address. A follow-up interview is required to determine if the census address is actually a duplicate of another census address.

RV = The match status of this A. C. E. or census address is not clear. A review by a Technician or Analyst is needed to resolve this case. (This is a temporarily assigned code that is resolved before the follow-up interview.)

There are three categories of matches, which are coded M:

- The A. C. E. and census addresses matched exactly.
- No address match, but the location of the map spot on the A. C. E. map and the census map refer to the same housing unit or the location descriptions match.
- The basic address matches for a multi-unit building or a mobile home park. The unit descriptions did not match, but there is evidence the A. C. E. units and the census units referred to the same unit (i.e., the unit descriptions for the multi-unit or mobile home park are non-contradictory).

There is a separate field in the database to link possible duplicates for both the A. C. E. and the census. In before follow-up housing unit matching, the primaries to be linked to duplicates are housing units with codes of M, NE, NI, and RV. The housing units coded DE, DI, and P are not allowable primaries.

### 2.4.3 Technician Review

The Technicians will perform two activities in the housing unit matching: review the difficult cases and perform the quality assurance for the clerical matching. The difficult cases are the ones that the clerical matchers could not code (i.e., RV) or ones where they see something unusual. The clerical matchers have been instructed to use a code that flags the cases for a higher review. This speeds up the matching and increases the efficiency of the matching. When an RV code is entered for a case by the clerical matcher, the case is automatically flagged for review by the Technicians. In addition to clusters with RV codes, they will review clusters with address changes or follow-up notes,

The Technicians will also be responsible for the quality assurance for the housing unit matching. All of the work done by the clerical matchers will be reviewed initially until the clerical matcher is determined to be performing at an acceptable level of quality. The number of records to be reviewed before the clerical matcher is classified as acceptable is 200. After a clerical matcher's work is acceptable, a systematic sample of clusters completed by each clerical matcher will be reviewed for quality assurance. The software will continue to assess the level of quality of the clerk's work. If the clerk's work in the sample of reviewed clusters falls below the acceptable level of quality, the clerk will return to having all of their work reviewed by technicians.

If there are address records created in A. C. E. that were the result of keying errors, the clerks code the case with an RV and the technicians will use the ZI code to remove them from processing.

ZI = The A. C. E. address is incorrectly included in the A. C. E. list of housing units. This error is identified clerically after the A. C. E. list is created and does not need to be sent to the field for an interview. This code removes the address from further processing.

The technicians will not be able to correct A. C. E. map spot numbers for the 2000 A. C. E. In the 1995 and 1996 tests, the errors in map spot numbering were corrected by deleting the listing for the one with an error and adding the address with the correct map spot number. These map spot errors are made by the lister or there was a keying error. The editing and quality assurance were increased for the keying for dress rehearsal, but there still were a few errors. Housing units with incorrect map spot numbers will be removed from the A. C. E. listing and the P-sample by entering the ZM code.

ZM = The map spot number associated with a housing unit is in error. A delete code of ZM is entered for that housing unit. This code removes the housing unit from further processing.

The matching using maps spotted maps is done on the computer. The paper census maps with spots indicating the location of census housing units will be destroyed. The inserts created in the field will be saved, since they will not be scanned. The technicians will request the paper copies of the map inserts and use them as needed.

### 2.4.4 Analyst Review

The Analysts will perform two activities in the housing unit matching: review the difficult cases and perform quality assurance for the Technicians. The difficult cases are the ones that the Technicians could not code or ones where they see something unusual. When an RV code is entered for a case by the technician, the case is automatically flagged for review by the Analysts.

In addition to clusters with RV codes, the Analysts will also review clusters with ZI or ZM codes and clusters where the Technician changed more than half of the clerks's codes.

The Analysts will also be responsible for the quality assurance for the housing unit matching. A systematic sample of clusters completed by each Technician will be reviewed for quality assurance after the Technician has been classified as qualified.

## 2.5 Clusters With No Clerical Matching

Clerical matching is labor intensive. We will reduce the amount of clerical work performed for the 2000 A. C. E. by identifying clusters by computer that can be sent for a follow-up interview without a clerical review. Each of the following four sets of criteria determines, in priority order, which clusters skip clerical review and are sent to field follow-up without clerical review. For all criteria, clusters must have no more than 4 housing units coded P.

- Criteria for all urban and rural clusters

  1. Census nonmatches, but no A. C. E. nonmatches
     Clusters must have zero housing units coded NI.
  2. A. C. E. nonmatches, but no census nonmatches
     Clusters must have zero housing units coded NE.

- Additional criteria for all urban areas and only rural areas that have 100 percent house number and street name addresses in both the A. C. E. and census universes.

  1. Clusters must meet all three additional conditions:
     a. The sum of P, NI, and NE (P+NI+NE) is less than or equal to fifteen.
     b. The sum of NI and NE (NI+NE) is not equal to zero.
     c. NI must not be equal to NE when NI is less than six

  2. Clusters must meet both additional conditions:
     a. The absolute value of the difference in NI and NE ( |NI-NE| ) is greater than or equal to eleven.
     b. The difference between the sum of P, NI, and NE and the absolute value of the difference in NI and NE is less than or equal to fifteen.

This criteria attempts to identify clusters where the payoff of clerical review is small. If the clerks do little in the clerical matching, we might as well send them to the field without clerical review. This way the clerical matchers can concentrate on the more difficult clusters where the review will be beneficial and will reduce follow-up. If the clerical review does not reduce follow-up, there is no reason to do it.

## 2.6 Housing Unit Follow-up

The cases coded as not matched, possibly matched, or possible duplicates are sent for follow-up interview for all types of basic street address codes. Selected matched cases are also sent for additional information. Specifically, the cases identified for field follow-up are:

- The A. C. E. addresses with a before follow-up code of NI. These A. C. E. housing units are map spotted and information is obtained to confirm if they were housing units within the block cluster.
- The census addresses with a before follow-up code of NE. Information is obtained to confirm if these housing units were housing units within the sample block cluster.
- The possible matches are sent to the field to determine if the A. C. E. and census addresses refer to the same housing unit. If they do not match, then they are interviewed as an A. C. E. nonmatch and a census nonmatch during the housing unit follow-up.
- Census housing units that are identified as possible duplicates are followed up to determine if the two census addresses refer to the same housing unit.
- A. C. E. housing units that are identified as possible duplicates are followed up to determine if the two A. C. E. addresses refer to the same housing unit.
- Matched housing units with a unit status code of 2, 3, 4, 7, and 8 (2 = under construction, 3 = future construction, 4 = unfit for habitation, 7 = vacant trailer site in a mobile home park, 8 = other).

The A. C. E. housing unit with unit status indicating something other than an occupied or vacant housing unit that is intended for occupancy need a follow-up interview to determine their status at the time of the follow-up interview.[5] The address will be either classified as a housing unit or the address will be removed from further processing. For example, a unit that is under construction or future construction at the time of listing may fit the definition of a housing unit at the time of the follow-up interview. If the unit fits the definition of a housing unit, it is included in the A. C. E. housing unit processing. If construction has not progressed enough to fit the definition of a housing unit, it is coded as removed from the A. C. E. housing unit inventory.

The housing unit follow-up forms are computer generated. The housing units requiring a follow-up interview and the questions are printed. In addition, all housing units in the block cluster are printed for reference. This list of addresses in the block cluster is known as the housing unit reference list. Census addresses are on one side of the form and A. C. E. addresses on the other side. The housing units within a block cluster are printed in A. C. E. map spot and within map

[5] Housing units classified as vacant and boarded up or housing units used to store household goods fit the census definition of a housing unit, but are not occupied. These units matching the census are not followed up because they fit the census definition of a housing unit. See Section 6.8 for more information on census definitions of housing units.

spot number identification number order within a block. Census nonmatches are printed within the cluster in the proper order for city style addresses. In the rural areas, the census nonmatches are inserted in the listings using the census map spot numbers.

The objective of the follow-up interview is to create an accurate listing of all housing units in the block cluster and to link different versions of the housing unit address that could appear on the independent listing and census lists. This revised listing of housing units confirmed to be in the block cluster after the housing unit follow-up is called the preliminary enhanced list[6]. The addresses in the census that were not originally listed in the A. C. E. listing book are not eligible to be included in the P-sample and are not included in the person interviewing.

The questions on the follow-up form are not designed to be read to respondents, but are used as a guide for an interviewer. The answer to one question may be the result of asking several other questions. The interviewer appropriately modifies the questions, when necessary, to the situation that is encountered in the field and records the appropriate answers on the follow-up form.

For example, the interviewer is to determine if the address for the Independent Listing or census nonmatch existed as a housing unit. This is not a question for a respondent. There are several reasons why an address might not fit the definition of a housing unit, such as it burned, it was a mobile home that moved, it was converted to fewer housing units, it was group quarters, it was used for storage of farm machinery, it was the laundry room in an apartment complex, it was a business, and so forth. The interviewer appropriately modifies the questions, as necessary, to the situation that is encountered in the field.

Corrections and updates to the addresses are also recorded on the follow-up form. The address updates are keyed into the database to accurately identify housing units for the person interviewing. The follow-up interviewer will not add housing units missed by both the A. C. E. and census for the 2000 A. C. E.

## 2.7 After Follow-up Coding

### 2.7.1 Clerical Coding

After the field follow-up, the completed forms are returned to the processing office. Using the information obtained during the field work, an After Follow-up Match Code is assigned for cases sent to the field. These match codes are:

| | | |
|---|---|---|
| M | = | The A. C. E. and census addresses match. |
| CI | = | The A. C. E. housing unit existed as a housing unit at the time of the follow-up interview and is correctly geocoded in the block cluster. The housing unit is not found in the census. |

[6] See Section 2.9.2 for more on the preliminary enhanced list.

CE = The census housing unit existed as a housing unit at the time of the follow-up interview and is correctly geocoded in the block cluster. The housing unit is not found in the A. C. E.

ZI = The A. C. E. address did not refer to a housing unit at the time of the follow-up interview. The A. C. E. housing unit is removed from the enhanced list. For example, the housing unit burned, the mobile home moved, the address is commercial property, or a special place at the time of the follow-up interview.

EE = The census housing unit is erroneously listed on the DMAF, because the address is not a housing unit at the time of the follow-up interview in the block cluster. For example, the housing unit burned or the mobile home moved. Another example, the address is commercial property or a special place. Also, the address is nonexistent within the sample block cluster.

GI = The A. C. E. housing unit existed as a housing unit at the time of the follow-up interview, but is incorrectly listed in the block cluster. The housing unit is an A. C. E. geocoding error.

GE = The census housing unit existed as a housing unit at the time of the follow-up interview, but is incorrectly geocoded to this block cluster. This housing unit is erroneously enumerated in this block cluster, because of a geocoding error.

DE = The census housing unit is erroneously enumerated in the census. The reason for erroneous enumeration is the address is duplicated in the census.

DI = The housing unit should not have been listed in the A. C. E. This address is a duplicate of another A. C. E. address. This address is removed from further processing for the A. C. E. and is not included on the enhanced list.

MU = The A. C. E. and census addresses match and there is not enough information on the follow-up form to confirm this match as a housing unit with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.

UI = Not enough information on the follow-up form to assign a code to the nonmatched A. C. E. housing unit with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.

UE = Not enough information on the follow-up form to assign a code to the census nonmatched housing unit with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.

RV = The match status of this A. C. E. or census housing unit is not clear. A review by either a Technician or Analyst is needed to resolve this case. (This is a temporarily assigned code that is resolved before the after follow-up coding is completed.)

The reference to "on census day" in codes that record the results of the follow-up interview, which are CI, CE, ZI, EE, GI, and GE, have been replaced with "at the time of the follow-up interview". The A. C. E. design originally was to construct a list of addresses confirmed to exist as housing units on census day. A. C. E. person interviews are conducted at these addresses. A telephone interview phase was added in order to reduce the interviewing burden. The timing of the housing unit follow-up was moved to start before census day. Therefore, all references to census day status were removed.

If two housing units identified as possible duplicates are in fact duplicates, the housing unit is coded as a confirmed duplicate. Neither A. C. E. nor census duplicates are included in the enhanced listing for person interviewing. If follow-up determined they are not actually duplicates, the duplicate code and linking information are removed. In after follow-up housing unit matching, the primaries to be linked to census duplicates are census housing units with codes of M, MU, CE, UE, and RV. The housing units coded DE, EE, and GE are not allowable primaries. The A. C. E. housing units available as primaries are M, MU, CI, UI, and RV. The A. C. E. housing units coded DI, ZI, and GI are not allowable primaries.

Census housing units coded as CE or UE will be assigned an A. C. E. map spot number. They will be inserted where they belong in order with respect to the other housing units in the block. For example, if a census correct enumeration belongs between map spot 19 and 20, the census correct enumeration will be assigned map spot 19A. Therefore, all structures on the preliminary enhanced list are in the order in which they appear on the ground and the subsampling in large clusters will select contiguous segments of housing units for A. C. E. person interviewing. The map spot numbers are assigned in the field for the cases coded CE and in NPC for the cases coded UE.

### 2.7.2 Technician Review

The Technicians will review the clusters with the review code and perform a quality assurance for the clusters processed in the after follow-up housing unit matching by the clerical matchers. A systematic sample of clusters will be selected for quality assurance from each clerical matcher's work after they are classified as qualified for clerical matching.

The Technicians will be able to use the ZM code in after follow-up coding, since some clusters will be seen for the first time by clerks.

ZM = The map spot number associated with a housing unit is in error. A delete code of ZM is entered for that housing unit. This code removes the housing unit from further processing.

### 2.7.3 Analyst Review

The analysts will review the clusters with the review code and perform a quality assurance for the clusters processed in the after follow-up housing unit matching by the technicians. A systematic sample of clusters will be selected for quality assurance from each technician's work after they are classified as qualified to be a technician.

## 2.8 Relisting for Clusters with A. C. E. Geocoding Errors

There may be A. C. E. geocoding errors in the original A. C. E. housing unit listings. These housing units are coded GI during the after follow-up coding. If a large proportion of the A. C. E. housing units in the cluster are coded GI, the cluster will be relisted. The regional office will be notified and a blank listing book will be used to list the housing units in the cluster again. The field lister must be someone who has had no contact with this cluster before. The identification of clusters for relisting are identified when the after follow-up matching is completed. The decision to relist is automated. If 80 percent of the cluster is coded GI[7], the cluster is relisted.

The A. C. E. housing units must be collected independent of the census housing units. To have independence the A. C. E. housing unit listings must be done without the A. C. E. lister seeing the census inventory of housing units. If the A. C. E. lister sees the census housing units, independence is violated. We collect the A. C. E. housing units and match them to the census. If there exists a large number or proportion of housing units that do not match, there is a possibility of geocoding error for the A. C. E. or census housing units. If the A. C. E. housing units are incorrectly geocoded to this cluster, we must start over and relist the cluster for A. C. E. using a different lister. If the follow-up interviewer corrects the A. C. E. listings, independence is violated because the person collecting the A. C. E. listing of housing units has seen the census.

The number of relisted clusters is dependent on the quality of the TIGER maps, A. C. E. listing, quality assurance not detecting errors, and housing unit follow-up. It is impossible to predict the number of clusters needing relisting in 2000. The relisting may also not be uniform across regions. There is no maximum number of relisted clusters in any region.

There will be no housing unit matching in the relisted clusters during the housing unit matching phase of A. C. E. The addresses listed for A. C. E. during the relisting operation will be the addresses used to conduct person interviewing. These clusters will be treated in the same way as the list/enumerate clusters in 2000. All housing unit matching will be conducted for relisted and list/enumerate clusters during the final housing unit matching. Since there is no housing unit matching, there is no telephone interviewing in the relisted clusters. The phone number comes from the census questionnaire. All person interviewing in relisted clusters is by personal visit.

A match code of UI will be assigned to the A. C. E. housing units in the relisted clusters and the

[7] If GI / (GI + M + CI) is 80 percent or greater, the cluster will be relisted.

list/enumerate clusters. The census housing units in the relisted clusters will be assigned a code of UE and will be included in the preliminary enhanced list. See Section 2.9.2 for more discussion of the preliminary enhanced list.

### 2.9 Reduction Within Large Block Clusters

The small block cluster subsampling occurs in late January through early February 2000 following the completion of the keying of the independent listing books. It's important to wait for the results of the keying operation to get the best available number of housing units in the cluster. The small block cluster subsampling operation overlaps with the start of the housing unit matching. These two operations can overlap at this time because only the small block clusters selected in the A. C. E. listing sample are eligible for this subsampling operation, and the housing unit matching can start with the other clusters. However, it's important to finish the small block cluster subsampling before the housing unit matching operation ends. Small block clusters must go to housing unit matching as well.

Before this operation the A. C. E. reduction sample contains 5,000 small clusters in the United States and 96 small clusters in Puerto Rico. Small clusters are expected to have between zero and two housing units based on the early census address list. Conducting interviewing and follow-up operations in clusters of this size are not as cost-effective as in larger clusters. Therefore, to allocate A. C. E. resources more efficiently, we will only include a subsample of these small clusters in the A. C. E. interviewing sample.

This subsampling operation will reduce the sample of small clusters while at the same time attempting to balance three goals. First, we would like to prevent any small clusters from having weights that are extremely high compared to other clusters in the sample. Second, we would like to have lower weights on clusters where the number of housing units is different than we expected. These first two goals attempt to reduce the contribution of small clusters to the variance of the dual system estimates. The third goal is to ensure that the field staff can efficiently manage the resulting workloads.

Using the keyed independent listing counts and the January DMAF count, the small block clusters within each state are stratified by size. Then, a systematic subsample is selected from each stratum with equal probability. All small block clusters which have 10 or more housing units, are List Enumerate, or are on an American Indian Reservation or other type of American Indian Country are kept in the A. C. E. sample.

Approximately 1500 of the 5000 small block clusters are expected to be retained in the A. C. E. sample. This is the last stage of cluster reduction; therefore, there is expected to be approximately 12,500 block clusters in the A. C. E. sample for the 50 states and the District of Columbia. For Puerto Rico, there will be roughly 500 block clusters.

### 2.9.1 P-Sample

Following the completion of the housing unit matching and follow-up operations, the large block cluster subsampling operation can begin. This is continual operation, done on a cluster-by-cluster basis. Any block cluster with 80 or more independently listed housing units after the matching and follow-up operation is eligible to be subsampled. After all large block cluster subsampling is done, the interview sample size for the fifty states and the District of Columbia will be roughly 300,000 housing units. This stage of subsampling reduces the interviewing workload by reducing the number of housing units in a cluster to be in sample. The A. C. E. interview sample is the P-sample.

Large block cluster subsampling is done by forming segments of adjacent housing units and selecting one or more segments for A. C. E. person interviewing. The segments are of approximately equal numbers of housing units within a block cluster. The block cluster will be subsampled for the P-sample when it contains 80 or more A. C. E. housing units. That is, the decision to subsample is based on the number of addresses listed in the A. C. E. and confirmed to exist within the block cluster.

This subsampling is conducted in an automated environment. The results of the segmenting and subsampling are an important input to the E-sample identification.

### 2.9.2 E-Sample

The housing unit matching and follow-up is important for obtaining overlapping E- and P-samples. One of the key pieces of information linked to the independently listed housing units during these operations is the census identification number of the corresponding census address. The source of the E-sample is the Census Unedited File (CUF) which will be available in the fall of 2000. All census records in the A. C. E. block clusters are eligible to be in the E-sample. However, to reduce the person follow-up workload, a subsample of the census housing units is identified when there are more than 80 census housing units in the block cluster.

To create overlapping E and P-samples within a cluster which needs to be subsampled, the P-sample large block cluster subsampling results are mapped onto the census records on the CUF in the block cluster. This is possible because when there was a correspondence between an A. C. E. independently listed address and a census address during the initial housing unit matching, the census identification number was assigned to the A. C. E. unit. This facilitates mapping the large block cluster segmenting and subsampling. If there are a significant number of census addresses which do not correspond to an A. C. E. unit, then a further subsampling will be done to reduce the E-sample person follow-up workloads.

There are several instances where the E- and P-samples may not overlap within a cluster if there are more than 80 census housing units in the cluster. This usually occurs when it is operationally infeasible to assign a census identification number to an A. C. E. independently listed housing unit during the housing unit matching. One example is List Enumerate clusters. Since the List Enumerate census operation will not be completed by the time of the housing unit matching, there will not be any census address information available to link the A. C. E. and census addresses.

### 2.9.3 The Preliminary Enhanced List

The preliminary enhanced list contains all housing units confirmed to exist in the block cluster. The following types of housing units are eligible to be included on the preliminary enhanced list:

- A. C. E. and census housing units that are matched.
- A. C. E. housing units that did not match to the census, but are confirmed during follow-up to exist within the block cluster.
- Census housing units that are not in the A. C. E., but were confirmed during follow-up to exist within the block cluster.
- A. C. E. housing units with unresolved status.
- Census housing units with unresolved enumeration status.

The following housing unit codes are on the preliminary enhanced list: M, MU, UI, UE, CI, and CE. Any A. C. E. and census housing units without enough information from the follow-up interview are unresolved. These unresolved housing units are included on the preliminary enhanced list.

The objective of the follow-up interview was to create an accurate listing of all housing units in the block cluster and to link different versions of the housing unit address that could appear on the independent listing and census lists. This updated listing is used as the inventory of housing units for subsampling large blocks. This revised listing of housing units confirmed to be in the block cluster after the housing unit follow-up is called the preliminary enhanced list. The addresses in the census that were not originally listed in the A. C. E. listing book are not eligible to be included in the P-sample.

After the housing unit matching is completed, all housing units on the preliminary enhanced list are ordered by A. C. E. map spot number. This facilitates the subsampling. Any census housing units that did not match the A. C. E., but have been confirmed to be housing units in the block cluster after the original listing, were inserted in their proper location by adding a letter suffix to the map spot number. The individual units in multi-unit basic street address are ordered by apartment number in the A. C. E. listing books, but the added census housing units are inserted at the end of the list of units in the multi-unit. For example, apartment 102 is a census nonmatch and was confirmed to have existed as a housing unit. It will not be inserted between 101 and 103, which are A. C. E. housing units. Instead it will be inserted at the end of the list of A. C. E. multi-units, which may have been after apartment 1215.

### 2.9.4 The Subsampled Preliminary Enhanced List

The subsampled preliminary enhanced list contains all of the information in the preliminary enhanced list plus the information from large block cluster subsampling which indicates whether or not a housing unit is in the P-sample. This file is used for E-sample identification and creation of the enhanced list.

### 2.9.5 The Enhanced List

The enhanced list is a list of all of the housing units on the subsampled preliminary enhanced list that are to be interviewed during person interviewing. The enhanced list does not contain census housing units not in the list of A. C. E. housing units. These census housing units are coded CE or UE. In previous tests, the housing units on the subsampled preliminary enhanced list that are in the census and not in the A. C. E. (i.e., CE and UE) were interviewed to reduce the person follow-up workload for E-sample people who do not match to the P-sample. For 2000, only the P-sample housing units will be interviewed, but the subsampled preliminary enhanced list will still be created to aid in E-sample identification. Clusters with a high proportion of A. C. E. housing units with geocoding errors were relisted. The relisted A. C. E. housing units will be included in the enhanced list. If these clusters are large, they will be subsampled. The enhanced list contains only the P-sample housing units and is the list of addresses interviewed in the person interview. See Section 4.2 for a description of the P-sample identification.

## 3.0 The Interviewing Phase

### 3.1 Person Interview

The A. C. E. person interview is conducted using a CAPI instrument. In order to get an early start for the interviewing, a telephone interview will be conducted for households where the census questionnaire is data captured and included a telephone number. Both mail returns and enumerator filled questionnaires are eligible for telephone interviews. Housing units without house number and street name addresses and housing units in small multi-unit structures will be excluded from the telephone interviewing. Large multi-units are included in the telephone interviewing because they tend to have unique unit designations. Many small multi-unit structures and rural areas do not have addresses that allow the telephone interviewer to distinctly identify the address. All remaining interviews after the telephone operation is completed will be conducted in person, but some nonresponse conversion operation (NRCO) interviews and interviews in gated communities or secured buildings may be conducted by telephone.

The person interview is conducted only with a household member for the first three weeks of interviewing. If the interview with a household member is not successful after three weeks, an interview with a nonhousehold member is attempted, called a proxy interview. The proxy interviewing is allowed during the remainder of the interviewing period. During the last two weeks of interviewing a nonresponse conversion operation is attempted for the noninterviews using the interviewers considered to be the best available. This noninterview conversion will attempt to obtain an interview with a household member or a proxy respondent, but not a last resort interview. This noninterview conversion will reduce the noninterview rate.

There are three paths or sections within the person interview. An interview is conducted using the first two paths when at least one person lives at the housing unit being interviewed. One path collects data from a household member and another path collects data from a nonhousehold member (i.e., proxy respondent) for these people. There are two paths because the questions are worded differently for interviews with household members and with proxy respondents. The interviews from the first two paths are in housing units containing whole household nonmovers, whole household inmovers, or households with a mixture of nonmovers, inmovers, and outmovers.

The third path is for whole household outmovers. The data for outmovers is obtained by proxy with the current resident in the sample household or with other proxy respondents when necessary.

After the names and characteristics are obtained, the resident status on census day is established. For nonmovers and outmovers, mover status in addition to questions about group quarters and other residences on census day establish the residence status[8].

In 1995 and 1996, the housing unit follow-up interview was conducted after census day and the status of the housing unit on census day was established during the interview. The follow-up interview is scheduled to begin before census day for the 2000 A. C. E. Therefore, the wording of the questions was changed in dress rehearsal to establish the status of the housing unit at the time of the follow-up interview. Since the follow-up in the housing unit phase did not identify the housing units that did not exist on census day, the A. C. E. person interviewers will identify these housing units and code them as nonexistent on census day.

The interviewers are not allowed to add housing units during the A. C. E. person interview, because people obtained by adding housing units to the enhanced listing are not in the P-sample for Dual System Estimation. Also, the option to add housing units was used incorrectly in 1995 and 1996. Address corrections were made by coding the housing unit as nonexistent and adding the housing unit with a corrected address instead of correcting the address. The housing unit was removed in error, resulting in the housing unit being removed from the P-sample.

The interviewers are allowed to make changes to the address during the person interview. These changes should only be updates and changes that help uniquely identify the housing unit. For example, the interviewer should add house numbers when no house number was listed in the A. C. E. address. The interviewer is not to make changes such as correcting 1101 Hollyhills Rd to 1101 Holly Hills Rd.

[8] See Section 6.9 for census residence rules.

### 3.1.1 Interviewing Outcome Codes

The interviewing outcome codes are assigned within the person interview to identify complete and partial interviews, the types of noninterviews, vacant housing units, and units that did not exist as housing units.

**Outcome 200** - Unopened Case - This is a new case that has not been opened or is a case that has been sent to supervisory review and has been reassigned and not opened by the new interviewer. This outcome will never leave the laptop with the initialized action code of 00. The interviewer must take some action to resolve it.

**Outcome 201**- Complete Interview - This is a completed interview. We have name, age and sex (no "don't know" or "refused") for each person in the current household. We have no "don't know" or "refused" answers for any of the probes for additional people or additional residences. None of the members of the household have an unresolved residence status code (i.e. we know if they lived there on census day).

**Outcome 202** - Insufficient Partial - This case has been opened but does not have sufficient information to qualify for any other outcome. This case should not leave the laptop.

Note: Field uses the terms Type A, Type B, and Type C to categorize noninterviews for reports and interviewer evaluation. A Type A is a noninterview that an interviewer theoretically should have obtained. This category typically includes refusal, temporarily absent, no one home, an various other types of non-contacts with an occupied unit. The other types of noninterviews (Type B or C) are cases that are temporarily out of scope, for example a vacant home, or Type B, or permanently out of scope, i.e., a demolished home, or a Type C.

**Outcome 203** - Partial Interview - This is a partial interview. We know the names of the current residents and have answers for age, sex and the probes, but the answers can be "don't know" or "refused". Residence status may be unresolved.

**Outcome 213** - Type A noninterview - We are unable to complete this interview due to a language problem with the current residents of the sample unit.

**Outcome 216** - Type A noninterview - We are unable to complete this interview due to our inability to find the current occupants of the unit or a knowledgeable respondent for the current status.

**Outcome 218** - Type A noninterview - We are unable to complete this interview due to a refusal by the current occupants of the sample unit.

**Outcome 326** - Type B noninterview - This unit is currently vacant.

**Outcome 333**- Type C noninterview - This unit currently does not exist or currently is not a housing unit.

### 3.1.2 Edit for CAPI Data Review

The CAPI data review is an opportunity to repair items that fail edit in the person interview data. These items include the outcome code, respondent code, A. C. E. status code, and any unexpected errors in the instrument that may occur. After the data is returned to headquarters a computer edit is used to recode blank or invalid codes in outcome, respondent, or A. C. E. status code. The data is also edited for households containing only people under 16 years of age. Another edit is to not allow households with inmovers to have unresolved residence status. In addition there is some recoding for QA outcomes that are replacement interviews.

### 3.1.3 A. C. E. Mover and Residence Status Codes

In 1995 and 1996, the CAPI instrument identified people who lived at the sample address on census day. Each A. C. E. person was assigned a residence status code of resident or nonresident of the housing unit on census day. If the residence status could not be determined from the information collected in the interview, the residence status was unresolved.

With procedure C, each A. C. E. person is assigned an A. C. E. mover code, A. C. E. born since census day code, and an A. C. E. status, which is the mover status and residence status combined. Procedure C was used in the Dress Rehearsal Accuracy and Coverage Evaluation and will be used in the 2000 Accuracy and Coverage Evaluation. See Section 6.1 for more on Procedure C.

- **A. C. E. Mover Code**

| | | |
|---|---|---|
| 1 | = | Nonmover |
| 2 | = | Inmover |
| 3 | = | Outmover |

- **A. C. E. Born Since Census Day Code**

| | | |
|---|---|---|
| 0 or blank | = | Default for inmovers |
| 1 | = | Not born since census day |
| 2 | = | Born since census day |
| D | = | Don't know |
| R | = | Refused |

- **A. C. E. Group Quarters Code**

| | | |
|---|---|---|
| 0 or blank | = | Default for inmovers |
| 1 | = | In group quarters on census day |
| 2 | = | Not in group quarters on census day |
| D | = | Don't know |
| R | = | Refused |

- **A. C. E. Other Residence Code**

| | | |
|---|---|---|
| 0 or blank | = | Default for inmovers |
| 1 | = | In other residence on census day |
| 2 | = | Not in other residence on census day |
| D | = | Don't know |
| R | = | Refused |

- **A. C. E. Status**[9]

| | | |
|---|---|---|
| N | = | Nonmover, resident on census day |
| O | = | Outmover, resident on census day |
| I | = | Inmover, nonresident on census day |
| R | = | Removed, nonresident on census day |
| U | = | Unresolved residence status |
| B | = | Born since census day, nonresident on census day |

The A. C. E. status code will be on the screens for the clerical matching. The nonmovers are people who live at the interviewed housing unit on census day and at the time of the A. C. E. person interview. The outmovers are people who lived at the interviewed housing unit on census day, but moved to another address since census day. The nonmovers and outmovers were both residents of the housing unit on census day (i.e., the A. C. E. status is N or O).

The inmovers lived at the housing unit at the time of the A. C. E. person interview, but did not live in the housing unit on census day (i.e., the A. C. E. status is I). They were not residents of the housing unit on census day. The persons included in the person interview, but classified as removed were collected as nonmovers or outmovers, but should not have been counted in the housing unit on census day according to census residence rules (i.e., the A. C. E. status is R). They were identified as nonresidents because they were living in group quarters on census day or had another residence where they should have been counted on census day according to census residence rules. Both the inmovers and removed people were not residents of the housing unit on census day. Any person who was born since census day has an A. C. E. status of B.

People with unresolved residence status were collected in the person interview as nonmovers or outmovers, but not enough information was collected to resolve their residence status. The responses to the group quarters or other residence questions were don't know or refused. The A. C. E. status is U for nonmovers and outmovers with unresolved residence status.

## 3.2 A. C. E. Form Selection

The CAPI data for the last interview when there are multiple interviews for the same housing unit is the one selected for further processing. If there is also a QA interview that replaces the

[9] Some people call this the residence status, but it really is the mover and residence status combined and referred to in this document as A. C. E. status.

original interview, the QA interview is selected over any other interviews.

## 4.0 The Person Phase

The P-sample people and the census people from the Census Unedited File (CUF) are computer matched within cluster. The possible matches, P-sample nonmatches, and E-sample nonmatches are clerically reviewed using an automated computer match and review system. Additional matches and possible matches are identified by the clerical staff. Duplicates on both lists are also identified clerically. People with incomplete names are identified, because they do not contain sufficient information for matching and follow-up. After the matching is completed, field follow-up is conducted for selected cases and the results of the field interview are coded in the matching database.

P-sample people are people who were identified as nonmovers or outmovers and were residents of the A. C. E. housing unit on census day. People with unresolved residence status are included with the P-sample people to attempt to resolve their residence status during the A. C. E. person follow-up operation. Therefore, the P-sample people must have an A. C. E. status of N (nonmover), O (outmover), or U (unresolved residence status). The P-sample people must have been listed in a housing unit that was listed during the housing unit phase of the A. C. E. In addition, the A. C. E. person interview must be a complete or partial interview.

### 4.1 P-Sample Identification

The P-sample housing units are identified from the A. C. E. housing units on the subsampled preliminary enhanced list. The housing unit match codes assigned to the A. C. E. housing units are M, MU, CI, and UI. The census housing units with codes CE and UE are not in the P-sample. Interviews for A. C. E. are only conducted at P-sample housing units. All P-sample housing units are on the enhanced list.

The nonmovers, outmovers, and people with unresolved A. C. E. status (i.e., people with A. C. E. status of N, O, and U) in P-sample housing units are included in the matching for A. C. E. and are P-sample people for matching. The nonmovers and outmovers lived in the P-sample housing unit on census day. The people with unresolved residence status are included in the P-sample. Their residence status will be resolved during the person follow-up interview.

### 4.2 Preliminary P-sample Estimation Outcome Codes

Preliminary P-sample estimation outcome codes are assigned to each P-sample housing unit before the computer and clerical matching. This outcome code is the matching outcome code assigned to the housing unit as of census day for nonmovers and outmovers. Therefore, only people with A. C. E. status codes of N = nonmover resident, O = outmover resident, or U = unresolved residence status are used in the coding.

The preliminary P-sample estimation outcome codes are defined as follows:

1 = Complete interview with a household respondent
2 = Complete interview with a proxy respondent
3 = Partial interview (i.e., some, but not all P-sample people have sufficient information for matching and follow-up)
6 = Field noninterview
9 = No people have sufficient information for matching and follow-up
10 = No census day residents (All people are nonresidents.)
11 = Vacant on census day
12 = Not a housing unit on census day

P-sample estimation outcome codes 1, 2, and 3 are interviews, 6 and 9 are noninterviews, 10 and 11 are vacant, and 12 is not a housing unit within the block cluster. Only people in interviewed P-sample households (i.e., preliminary P-sample outcome codes of 1, 2, or 3) are eligible for matching.

When there are no people in a P-sample housing unit with sufficient information for matching and follow-up, any people found in that housing unit are not in the P-sample for matching and the estimation outcome code is converted to a noninterview (i.e., outcome code 9). People are not included in the P-sample when they should have been counted in group quarters or had another residence where they should have been counted. These people are not residents of the housing units on census day. When all people in the housing unit are nonresidents, the housing unit is vacant on census day (i.e., outcome code 10). When there is a mixture of all nonresidents and insufficient information for matching and follow-up, the estimation outcome code is the same as whole household insufficient information for matching, which is a noninterview (i.e., outcome code 9).

## 4.3 E-Sample Identification

The E-sample identification is one of the activities in preparing the data files obtained from the census for A. C. E. person matching. The objective of the E-sample identification is to determine which people and housing units from the census are counted in the same block cluster or segments of block clusters that were selected for the P-sample. An overlapping P-sample and E-sample will reduce the E-sample follow-up workload. No census group quarters are included in the E-sample. The E-sample identification is accomplished in two steps:

- mapping block clusters or segments of block clusters from the housing units on the preliminary enhanced list into the census housing units enumerated in the census
- subsampling if the number of E-sample housing units is too large

Subsampling large block clusters for A. C. E. interviewing occurs when the block cluster contains 80 or more A. C. E. housing units. Large clusters on American Indian Reservations are not subsampled. Subsampling large block clusters is done by segmenting the block cluster into groups of contiguous housing units and selecting a sample of segments. Subsampling to generate the A. C. E. enhanced list for a cluster occurs before the A. C. E. interviewing begins. P-sample housing units are identified from the housing units listed for A. C. E. The P-sample people are identified from the A. C. E. person interview in P-sample housing units. When an A. C. E. address is matched to a census address, we consider that census housing unit as being on the preliminary enhanced list. Nonmatched and unresolved census housing units can also be on the preliminary enhanced list.

The housing units in the census are the housing units on the CUF. The CUF is used to identify E-sample housing units. The people on the CUF in these E-sample housing units are E-sample people. Each census person enumerated during the census in sample block clusters is assigned an E-sample indicator and an E-sample probability code. The E-sample indicator and E-sample probability code are attached to all census person and housing unit files used in person matching.

**E-sample indicator**[10]

1 = The person enumerated in the census was enumerated in a housing unit that was selected to be included in the E-sample.

2 = The person enumerated in the census was enumerated in a housing unit that was not selected to be included in the E-sample.

It is possible for E-sample housing units in a block cluster to be sampled at different rates. Probabilities of selection will be maintained at the block cluster level and the E-sample probability code will link the housing unit and the people within the housing unit to the appropriate probability of selection on the design file[11].

**E-Sample probability code**

1 = Assigned to CUF housing units that meet one of the following criteria:

- in a block cluster where the total number of CUF housing units in the block cluster is fewer than 80, whether or not the housing unit was on the preliminary enhanced list.
- in a block cluster where the total number of CUF housing units in

[10] There is an E-sample indicator of 3, which is described in Section 4.7, Targeted Extended Search. E-sample indicators 1 and 2 are in the sample block cluster and 3 is in the surrounding blocks for clusters included in the targeted extended search.

[11] See DSSD Census 2000 Procedures and Operations Memorandum Series R-4, "Accuracy and Coverage Evaluation Survey: Sample Summary File and Sample Design File Documentation", prepared by Randy ZuWallack for more information.

the block cluster is 80 or more and the housing unit was on the preliminary enhanced list.

2 = Assigned to CUF housing units that are in a block cluster where the total number of CUF housing units in the block cluster is 80 or more and the housing unit was not on the preliminary enhanced list.

### 4.4 E-sample Insufficient Information for Matching and Follow-up

First, a quote from a paper presented at the 1980 Annual Meetings of the American Statistical Association in Houston, Texas by Howard Hogan and Charles D. Cowan, U.S. Bureau of the Census titled "Imputations, Response Errors, and Matching in Dual System Estimation".

> "The important constraint in designing a survey to measure erroneous enumerations, or defining sufficient information for matching is that the independence of the census and the followup survey must not be compromised. We must be careful not to turn the dual record system into a double entry accounting system. Not being able to locate a record in the other system cannot be allowed to be grounds for defining it II[12] or EE[13].
>
> This may seem obvious, but there is a strong tendency to do just that. It is extremely easy to set up a matching procedure which first searches the other system in an attempt to find a match. If a match is found, the record is coded "Matched". If no record is found, the personal and address information is carefully examined, and it is often determined that there was not really enough information for matching, and the case is coded II instead of being considered a true nonmatch (out of one system). The errors introduced by this biased procedure can be of the same magnitude as the number of true misses.
>
> Checking for sufficient information is a simple clerical problem. Clear rules can be established for minimal information. Establishing optimal rules for sufficient information however, is one of the most important decisions to be made. Overly stringent rules lead to increased correlation bias, as well as increased variance by reducing the size of the usable sample. Overly loose rules of sufficient information are bound to lead to an increase in erroneous nonmatches and erroneous matches, and these in all probability will not balance out. This is not a trivial problem–even in the case of imputations."

The census person records are reviewed by both computer and clerically to identify people with insufficient information for matching and follow-up. Only people with sufficient information for matching and follow-up are allowed to be processed in the matching and follow-up interviewing phases of the person matching. The three types of insufficient information are:

[12] II is insufficient information for matching and follow-up.

[13] EE is erroneous enumeration.

- The census people are not data defined.
- The census people are data defined, but computer coded as insufficient information for matching and follow-up.
- The census people are computer coded as sufficient information, but converted clerically to insufficient information for matching and follow-up.

The first type of census people who are not data defined are not included in the E-sample. Only data defined people are included in the E-sample. These data defined people create person records in the census.

### 4.4.1 Census Data Defined

The term "data defined" is a term that has been used in the past at the Census Bureau to mean that a census person record has been created. The term "Total Persons" is the total number of people counted in the census at a census ID. The term "Selected Persons" is the data defined census people at a census ID. The difference is the people who are not data defined. These people have no census person record. A whole person imputation procedure is employed to create characteristic data in the census for these people.

Two characteristics are required to be data defined, where name counts as a characteristic. Name must have at least three characters in the first and last name together. The characteristics that are included in the counting are relationship, sex, race, Hispanic origin, and either age or year of birth[14]. Census records are created on the CUF for all data defined people. Anyone who is not data defined is a whole person imputation.

The count of census people who are whole person imputations must be identified separately from the other census people with insufficient information for matching, because it is a separate term for them in the Dual System Estimator[15]. The number of whole person imputations is subtracted from the census count within post-strata. The E-sample people who are data defined and have insufficient information for matching are included in the count of erroneous enumerations and are subtracted out with the erroneous enumerations in the Dual System Estimator.

The mail return census forms have been designed to collect characteristics for six people. Space

[14] Person one does not automatically have a relationship of head of household like it did in 1990 and the telephone number in item 2 on the mail return questionnaire does not count as a characteristic. The age and date of birth are examined together. If age is present, age/year of birth counts as a characteristic. If age is blank, but year of birth is present, then the age/year of birth counts as a characteristic. If age and year of birth are both blank, the age/date of birth does not count as a characteristic. The month and day of birth were used in dress rehearsal in the determination of counting the age/date of birth as a characteristic, but not in the 2000 Census.

[15] See Section 6.5.

is provided to collect names for people in households with seven to twelve people. The short form collects census names for people in large households. Space is provided on the mail return short form to collect names for persons seven through twelve. The large household follow-up operation attempts to obtain characteristics for these people by telephone. The mail return long form has a roster for the household.

The exception is the enumerator questionnaire used in nonresponse follow-up. There is space for five people and a continuation form will be used to obtain data for persons six and above in large households.

There was discussion about names in the long form roster for persons seven through twelve being used to create person records and being data defined. If this had been adopted, the seventh name on the long form roster would create the seventh person record, the eighth name would create the eighth person, etc. It has been decided to not create person records when the census has names and no characteristics in large households. The A. C. E. will not attempt to create additional census data defined people for these people with only name in large households. These people will become whole person imputations. The number of whole person imputations used in the Dual System Estimator will correspond to the counts used in the census.

### 4.4.2 Computer Coding of Insufficient Information for Matching and Follow-up

The A. C. E. requires a minimum of information for matching and follow-up. A person with this minimum of information or more has sufficient information for matching and follow-up. A person with less than the minimum has insufficient information for matching and follow-up. The minimum amount of data required for the data defined census people to have sufficient information for matching and follow-up is complete name and two characteristics.

If a data defined census person has a blank or incomplete name, that person has insufficient information for matching and follow-up. These census people are coded by computer before clerical matching. For example, the census data is a white male age 44. He is data defined and a census person record is created. He has no name and is coded as insufficient information for matching and follow-up. We suppressed these people from the clerical matching in dress rehearsal. We will not suppress them in 2000 because we have an image retrieval system allowing us to convert some of the insufficient information for matching and follow-up to sufficient information. In order to control clerical error in matching people with less than sufficient information, the matching software will not let the clerks match to the census person until the data found on the image is entered into the system and that new data makes the person suffioient information.

Complete name is defined as:

- first name[16], middle initial, and last name
- first name and last name
- first initial, middle initial, and last name

Currently only age and year of birth are used to determine if age is present when counting characteristics to determine if the person has enough data to be data defined in the census. The A. C. E. will use the same criteria. In other words, when the age and year of birth are both blank, month and day of birth are not considered. If this criteria changes for the determination of census data defined, the A. C. E. will not change. Only age and year of birth will be reviewed in determining sufficient information for matching for A. C. E.

### 4.4.3 Clerical Coding of Insufficient Information for Matching and Follow-up

There are cases where the name is not blank, but should be insufficient information for matching and follow-up. Census names like Mr. Doe, Donald Duck, white female, etc. will be coded insufficient information by the clerical matchers. The computer can not recognize names that are not real or are really incomplete names.

We will review the image of the census questionnaire for census people coded as insufficient information for matching and follow-up to obtain additional data that may convert them to sufficient information for matching and follow-up. We can also allow children with first names and no last names, but with an adult that has a first and last name to be converted to sufficient information. We can also look at the roster for the long forms to get names. These updates to the names will be captured into the matching software. The software will be programmed to decide if the person has sufficient information for matching and follow-up. When the person has enough information to be sufficient information for matching and follow-up, the software will convert the match code.

## 4.5 P-sample Insufficient Information for Matching and Follow-up

The P-sample is reviewed by the computer to identify people who have insufficient information for matching and follow-up. The P-sample rules for sufficient information for matching and follow-up are the same as the E-sample rules. Sufficient information for matching and follow-up is complete name and two characteristics. These cases identified by the computer will be suppressed from viewing by the clerical matchers to prevent errors in matching people who should not be converted to sufficient information for matching. The probability of matched will be imputed for the P-sample people coded as insufficient information for matching and follow-

[16] The minimum number of characters to be a name is two. This applies to the first name and to the last name. Therefore, there must be two characters in the first name and two characters in the last name.

up. They are treated as the other P-sample people with unresolved match status.

There were instances in past tests where the P-sample person was matched by the clerical matcher in error. For example, the name for the P-sample person is "Female Johnson", which does not have a complete name. We could not match with certainty or follow-up the nonmatch with a high success rate. If we allow the P-sample person to be viewed by the clerical matchers, they might match "Female Johnson" to Mary Johnson, if the address and other characteristics matched. There would be a tendency to match some people or to code insufficient information which is unresolved when they could not match. This causes a bias in the match rate. If "Female Johnson" cannot be matched, we must allow her to become a nonmatch by sending her to follow-up. There probably would not be a high success rate in locating "Female Johnson" to conduct an interview. A noninterview results in an unresolved code. We might even locate the wrong "Female Johnson" and interview someone different from our P-sample person. The best way to keep these errors from happening is to suppress the P-sample cases computer coded as insufficient information for matching.

## 4.6 Within Block Cluster Matching

With procedure C, the people from A. C. E. housing units who will be initially matched to the E-sample and non E-sample census enumerations are:

- the nonmovers and outmovers identified as residents (i.e., A. C. E. status equal to N and O)
- the people with unresolved residence status (i.e., A. C. E. status equal to U)

The matching within the sample block clusters is done by the computer matcher followed by an automated clerical review. The computer matching will match the nonmovers and outmovers to the E-sample in subsampled blocks and then allow matching to the non E-sample census enumerations. These non E-sample enumerations are census people in housing units that were not included in the E-sample after the subsampling of census housing units. The E-sample people have an E-sample indicator equal to 1. The census people in the sample block cluster not in the E-sample have an E-sample indicator equal to 2.

The A. C. E. Technicians do the quality assurance for the clerical matchers and resolve the cases flagged by the clerical matchers as needing further review. The A. C. E. Analysts do the quality assurance for the technicians and resolve the cases flagged by the technicians as needing further review.

### 4.6.1 Computer Matching

During computer matching, the P-sample is matched to the census. However, this matching is prioritized; first the P-sample is matched to the E-sample, then any leftover nonmatches from the P-sample are matched to the non E-sample people. The Statistical Research Division computer matcher is used. The matching occurs in two steps:

- Record Pair Ranking: The standardized names from the P-sample person and the census-side person are compared along with the person characteristics. A ranking score is assigned to each pair of people and the optimal pairs are identified.

- Determination of Match Cutoffs: The optimal pairs in the cluster are reviewed to determine the cutoffs for matches and nonmatches. All pairs above the match cutoff are identified as a match. All pairs between the match cutoff and nonmatch cutoff are identified as possible matches. All pairs below the nonmatch cutoff are nonmatched. Match cutoffs are assigned conservatively so there are virtually no false matches.

The codes assigned during the computer match are:

M = The P-sample and census people match.

P = The P-sample and census people are possible matches.

NP = The P-sample person with sufficient information for matching and follow-up is not matched to the census.

NE = The E-sample person with sufficient information for matching and follow-up is not matched to a P-sample person.

KI = Match not attempted for the P-sample person because the person has insufficient information for matching and follow-up. The name is blank or incomplete or the name is complete but the person has only one characteristic. This is a computer assigned code and these people are suppressed from view by the matchers.

KE = Match not attempted for the E-sample person. The name is blank or incomplete or the name is complete but the person has only one characteristic.

N2 = The non group quarters, non E-sample person with sufficient information for matching and follow-up is not matched to a P-sample person.

K2 = The non E-sample person has insufficient information for matching and follow-up.

Q2 = The non E-sample person enumerated in group quarters with sufficient information for matching and follow-up is not matched to a P-sample person.

N3 = The non group quarters census person in the surrounding block with sufficient information for matching and follow-up is not matched to a P-sample person.

K3 = The census person in the surrounding block has insufficient information for matching and follow-up.

Q3 = The census person enumerated in group quarters with sufficient information for matching and follow-up in a surrounding block is not matched to a P-sample person.

The codes N2, K2, and Q2 are discussed in Section 4.6.2 and the codes N3, K3, and Q3 are discussed in Section 4.7.1.

### 4.6.2 Clerical Coding

The goal of the matching is to produce the correct ratio of cases classified as omitted to those classified as included in the census. After the computer matching, P-sample and E-sample people who do not match are reviewed clerically. The clerical matchers use the MaRCS software to enter their codes.

Two population counts are included in the MaRCS software: census person records and total census people. The census person records are the people from all census questionnaires after primary selection algorithm processing for the census ID that had two or more characteristics (i.e., data defined). The census name is counted as a characteristic. For example, you could have had a name on the census questionnaire with one characteristic and a census person record is created. The number of census person records is the number of people listed in matching, since there are no suppressed census people in 2000 A. C. E.

The total census people is the final population count for the census ID. The primary selection algorithm combined selected people from all census returns for the ID into the final number of people to be counted for each ID. The census person records count identifies the number of census person records that are included in MaRCS for the census ID. The total census people count identified the total number of census people counted in the housing unit. The difference in the two numbers is the number of non-data defined census people who are whole person imputations.

Each cluster is processed separately. If there are clusters near to one another, there can be an overlap of search areas. A sample block in one cluster could be a surrounding block in another cluster or two clusters could share a surrounding block. The matchers will not know this is occurring, because each cluster's sample and surrounding blocks will be complete. There will be no borrowing from other clusters. If two clusters share blocks, these blocks will be in the database for each cluster.

The census form type should be available in the software for the matchers. The census form type for the census person is an indication of the quality of the data and is helpful in matching.

The codes assigned by the clerical matchers are[17]:

M = The P-sample and census people match.

[17] All codes are assumed to be for a P-sample or E-sample person with sufficient information for matching and follow-up unless stated as insufficient information for matching and follow-up.

P = The P-sample and census people are possible matches. Additional follow-up is needed.

NP = The P-sample person is not matched to a census person.

NE = The E-sample person is not matched to a P-sample person, additional follow-up is needed.

KP = Match not attempted for the P-sample person, because (1) the name is incomplete, such as "Mr. Jones", or (2) the name is not a valid name, such as "White Female" or "Donald Duck". This is a clerically assigned code.

KE = Match not attempted for the census person. The name is incomplete or not a valid name, such as "Child Jones", or "Mickey Mouse". This is a clerically assigned code.

DP = The P-sample person is a duplicate of another P-sample person.

DE = The E-sample person is a duplicate of another E-sample person. The non E-sample person in the block cluster (i.e., E-sample indicator 2) is a duplicate of the E-sample person. The E-sample person is a duplicate of a census person in a surrounding block (i.e., E-sample indicator 3).

RV = The match status of this P-sample or E-sample person is not clear. A review by either a technician or analyst is needed to resolve this case.

FE = The E-sample person is a dog, cat, or other animal that should not go to follow-up. The enumeration is fictitious.

GE = The E-sample person is erroneously enumerated in this block cluster, because the housing unit is a geocoding error.

NC = The P-sample nonmatch was found on the census roster. This person in a partial nonmatch household was not matched to the census because only name was collected in the census for this person in a large household and the census person was not data defined. No follow-up interview is necessary.

GS = The E-sample person is enumerated in a housing unit that exists in a surrounding block.

GC = The E-sample person is enumerated in a housing unit that exists in the sample cluster.

GU = The geographic work for the targeted extended search is unresolved. The field work was not done or the block number on the form was not in the surrounding blocks, in the block cluster, or on the map. It is not clear where the housing unit is located.

The review code (i.e., RV) is used by the clerical matchers when they are not sure of the correct code to use or if there is something unusual about the case. The review code reduces the time required for matching, because the clerical matchers do not need to wait for a decision on an unusual case. They simply code it with the review code and the work is routed to the next level of matching for review.

The clerks are allowed to code a census person with the KE code when they encounter census people with names such as "Mr. X", "White Female", "Mr. Smith", "Child Jones", "Mickey Mouse", etc.

P-sample people with names that do not refer to real people are coded KP. Examples are "White Female" or "Donald Duck". We will not be sending these cases to person follow-up. They will remain insufficient information for matching and follow-up and have an unresolved match status.

The NC code identifies people that are not matched and need no follow-up. We would not use the NC with a P-sample person with unresolved residence status. The NC will not go to follow-up even if the person interview was with a proxy.

There are people not included in the E-sample because their housing unit was identified as non E-sample. These large block clusters contained 80 or more housing units. These people were assigned an E-sample indicator of 2. They are available for matching, but no further processing is required if they do not match. Since everyone must have a code, they are assigned a code of N2. When the computer has assigned a match or possible match between a P-sample person and a non E-sample person and the clerical matcher needs to unlink the pair, the N2 code is used for the non E-sample person who does not match. The N2 is automatically coded by the software.

In addition, the non E-sample people in the sample block cluster with insufficient information for matching and follow-up have been assigned a code of K2. All people coded K2 have an E-sample indicator of 2. The K2 code is used to code insufficient information and follow-up for census people in housing units and group quarters. People with the K2 code are not required to be reviewed when reviewing images for the E-sample people coded KE. The matchers are allowed to review images for updates when they feel it is needed to accurately code a case. The images are not to be used to update data for P-sample people. Any non E-sample person who is identified clerically as having insufficient information for matching will remain coded N2 and should not be matched.

After the CUF is pulled, a record is created to identify people enumerated in housing units and people enumerated in group quarters. The census people enumerated in group quarters are included in the matching. Census people in institutional and noninstitutional group quarters in the sample block cluster have an E-sample indicator of 2 and a match code of Q2. There is no computer matching to census people enumerated in group quarters.

Any census people enumerated in group quarters and having insufficient information for matching and follow-up are assigned the K2 code.

Matches between P-sample people and non E-sample people in the cluster are coded as M. The matches to census people within the cluster and subsampled out of the large blocks are identified by matches to census people with an E-sample indicator of 2.

The FE code is discussed in Section 4.6.3. The GE code is discussed in Section 4.6.5.1. The NC code is discussed in Section 4.6.5.2. The GS, GC, and GU codes are discussed in Section 4.7.2.

### 4.6.3 Census Images

For Census 2000, all census forms will be scanned and the subsequent information interpreted using OCR (Optical Character Recognition). A form is identified by the census ID (MAF ID plus the form extension) or a processing ID. Images will only be available for housing units on the January 2000 DMAF. Images will not be available for census housing units added after January 2000. An address may return more than one form, including the following: original census form, Be Counted form, a foreign language form, and/or an SEQ (Simplified Enumerator Questionnaire) form. For block clusters included in the A. C. E., the scanned images need to be available for the clerical matchers, technicians, and analysts to use during person matching. Be Counted forms will not be available to use for viewing images, since they do not have a census ID associated with the form when data captured.

During A. C. E., person matching must occur as fast as possible to allow for maximum time for follow-up for a case and subsequent estimation activities. Using our data capture system, a computer nonmatch may occur because of the unknown quality of scanned images (e.g., an incorrect capture of a name). As a result, clerical matchers could reduce the number of unresolved or nonmatch cases if they had access to the scanned census form images. Using the scanned images, the matchers could readily discern differences or problems in the responses that the scanning equipment could not recognize. For A. C. E., images will be captured in "real time" and written to CDs. Images will only be available for the sample block clusters. No images will be available for the surrounding blocks.

The clerical matchers will review census people with insufficient information for matching and follow-up to obtain additional information that may allow them to be matched. These people are coded KE. We will not suppress any of the census people with insufficient information for matching. All review of census people with insufficient information for matching and follow-up will be done before the matching begins. They will update the census data in the matching software. The software can check to make sure the matchers do not match anything that does not have enough characteristics. This prevents the clerk from matching to census people who do not have sufficient information for matching and follow-up. The software will not allow them assign another code until there are two characteristics and a complete name. The definition of name for sufficient information for matching and follow-up is unchanged. The census people with updated information will be recoded as NE (i.e., census nonmatch) by the software. Then matching can begin and the matchers can match the P-sample person to the new people coded NE. Census people converted from KE to NE are treated the same as people who began the matching with a code of NE.

The matchers will also be able to review data for nonmatches when we suspect data capture errors and to capture corrected data for names and all characteristics. The characteristics are relationship to person 1, sex, age, Hispanic origin, and race.

The corrected data is used on the follow-up form, but not sent to estimation. The updated data will not be put back on the CUF. This updating is for matching in A. C. E. and for the follow-up form only. It is not used for anything else. The matchers will **NOT** be looking at people who are not data defined to see if there is more information on the census form to make them data defined. Therefore, we will **NOT** be creating people.

The matchers will not code P-sample people as not matched but found on the census form, which is the old L code from 1990. We planned to use a code for these people to reduce the follow-up workload, but viewing the images for all P-sample nonmatches would have been time consuming for the small amount of follow-up reduction.

There were a few cases in dress rehearsal with information on the census questionnaire that identified an E-sample person as a dog or cat. We should use the fictitious code (i.e., FE) for these cases to keep them from going to follow-up.

### 4.6.4 Duplicate Search Within Cluster

The search for duplicates is done clerically. The printouts used in 1990 for duplicate search are automated in 2000. Search routines in the 2000 MaRCS make the searches quicker and more accurate. Duplicates are linked in the matching system for later analysis.

Duplicates are identified in both the P-sample and E-sample people. A P-sample person duplicate is removed from the final P-sample. Whole households of P-sample duplicates are converted to noninterviews for Dual System Estimation. The A. C. E. interview was not a good interview when the whole household was duplicated. An E-sample person duplicate is an erroneous enumeration in the census.

Any E-sample nonmatch that is identified as a duplicate of another E-sample person (i.e., both census people have E-sample indicators equal to 1) is assigned a code of DE. The two E-sample people are linked in the MaRCS software. The E-sample person coded DE is the duplicate and the other E-sample person is the primary. The E-sample person with the code of DE can be linked to any other E-sample person, except another E-sample person coded DE, KE, FE, and GE. For example, there is a triplicate and the codes are M, DE, and DE. The M and the first DE are linked and the M and the second DE are linked. The two people coded DE cannot be linked.

The E-sample people with E-sample indicator equal to 1 are also compared to the non E-sample census people with E-sample indicator equal to 2. When duplicates are identified between a census person with E-sample indicator equal to 1 and another census person with E-sample indicator equal to 2, the number of times that the E-sample person is duplicated with non E-sample people in the cluster is needed. Therefore, if the number of times duplicated with a census person subsampled out of the E-sample is equal to 1, then the E-sample person gets half of an erroneous enumeration. If the number of times duplicated is 2, then the E-sample person gets two thirds of an erroneous enumeration. In other words, when the number of times duplicated is 1, the probability of erroneous enumeration is one half and when the number of times duplicated is 2, the probability of erroneous enumeration is two thirds. Therefore, the formula for the probability of erroneous enumeration when the number of times duplicated is equal to d, is 100*d/(d+1) or

$$Pr\,(EE) = 100 * d / (d + 1)$$

where

d = number of times duplicated with census people in the cluster with E-sample indicator of 2.

This assumes the E-sample person has been coded as correctly enumerated. If the E-sample person is coded unresolved, the final probability of erroneous enumeration includes an imputation for unresolved enumeration status. If the E-sample person is assigned a match code that indicates erroneous enumeration, the number of times that the E-sample person is duplicated with non E-sample people is irrelevant and ignored. A person can not have a probability of erroneous enumeration that is larger than 100 percent.

We simplified the clerical coding by assigning all duplicates with the DE code. Then using the linking of duplicates between E-sample indicators of 1 and 2, the probability could be calculated. Therefore, all duplicates are coded DE for duplicates between two E-sample people and between an E-sample person and a non E-sample person. All duplicates are linked. When the E-sample person is duplicated with a non E-sample person in the sample block, the non E-sample person is assigned a code of DE. MaRCS calculates the number of times duplicated. The first time an E-sample person is duplicated with a non E-sample person, the number of times duplicated is calculated as 1. The second time the E-sample person is duplicated with a different non E-sample person in the sample block cluster, the number of times duplicated is changed to 2 for the E-sample person.

No duplicate search is conducted between the non E-sample people. Therefore, there can be no links between census people with E-sample indicators equal to 2. The number of times duplicated is assigned to the E-sample person. There is no duplicate search for census people counted in group quarters. Census people in group quarters should be suppressed when conducting clerical duplicate search.

### 4.6.5 Additional Clerical Coding

#### 4.6.5.1 Census Geocoding Errors

The clerical matchers review people in census housing units identified in the housing unit matching as geocoding errors. These housing units were coded GE during housing unit matching. The E-sample people in census housing units that are geocoding errors are erroneously enumerated. Therefore, the clerical matchers assign a code indicating geocoding error to E-sample persons.

The GE code is used only for whole household E-sample nonmatches. There is no need for a follow-up interview, since the housing unit follow-up operation identified these housing units with geocoding errors. These E-sample people are erroneously enumerated because they are enumerated in a housing unit that is a geocoding error.

### 4.6.5.2 Coding Nonmatches in Large Households

We will use the names on the rosters to reduce the P-sample follow-up of nonmatches. These people are still not matched to the census. We do not need to follow them up, because we know they are not matched and were residents of the housing unit on census day. They are not matched because they were in a large household and the large household follow-up was not successful.

The clerical matchers will review the household rosters for large households when one or more of the people have been matched. Only the nonmatches in partial household nonmatches will use the large household rosters to reduce follow-up. Therefore, when at least one person is matched of persons one through six on a census questionnaire, the extended roster on the short form can be used to code P-sample nonmatches as NC. This code indicates the P-sample nonmatch is a partial household nonmatch and does not require follow-up. Partial household nonmatches were 24.1 percent of the P-sample nonmatches in Chicago in 1996. The entire long form roster is used to identify partial household nonmatches in large households that will not be followed up because the name is included on the long form roster.

The term "data defined" is a term that has been used in the past at the Census Bureau to mean that a census person record has been created. If a "person" does not have enough information to be data defined, it may still influence the count if that person is represented in the variables used to determine household size. The difference between the data defined persons and the unit's household size is, in most cases, the number of whole person imputations for the unit. In 1990, a census person needed two 100 percent characteristics present on the questionnaire to be data defined. Name did not count toward data definition in 1990, because all names were not captured. Two characteristics were required in 1990 to help guard against "persons" being erroneously created by the scanning capture system due to a single mark in the 100 percent area of the questionnaire.

In 1995 and 1996, a person needed one characteristic to be data defined, since the data from the test census questionnaires were keyed. Name counted in 1995 and 1996 as a characteristic. Therefore, a person record could be created when only the name was present on the census questionnaire.

The census data will be captured by a scanning technique known as imaging. Optical character recognition (OCR) will be used to interpret the names with keying from image used as a backup if the names cannot be read with sufficient confidence by OCR. Person records are created for data defined census people when there are two characteristics captured from the census questionnaire with name counting as a characteristic. Whole person imputation will still be used for housing units whose determined size is larger than its data-defined person count. This count of imputed people is a term in the dual system estimator for A. C. E.

The mail return short form has a continuation roster to collect names for persons seven through twelve. The mail return long form has a roster for the names of persons one through twelve. Data were collected for the first six people in the household for both long and short forms. If the large household follow-up is unsuccessful, there will be only names for persons seven through twelve for the short form and only names in the household roster for persons seven through twelve for the long form. Person records will not be created for the people in large households with only names, since they are not data defined.

The A. C. E. will not be creating census person records to use in the E-sample for person matching, because the count of whole person imputations in the census is a term in the dual system estimator. In addition, the long form roster will not be used in large households for persons one through six to create additional E-sample people. Therefore, we will not be making any changes to the count of whole person imputations used in the dual system estimator by including the rosters for the long and short forms in the A. C. E. process.

The data defined persons are further reviewed for A. C. E. to identify people without complete names. Census people who do not have complete names are coded as having insufficient information for matching and follow-up. Some people may complete the household roster and leave the name blank on the person records. Guidelines need to be developed on when to use the household roster on the long form to make data defined census people with blank or incomplete names have sufficient information for matching and follow-up.

### 4.7 Targeted Extended Search

The targeted extended search for 2000 A. C. E. is a two stage process. First, clusters are identified that will benefit most from expanding the search area to surrounding blocks. Second, blocks within the cluster will be targeted for searching.

There are geocoding errors of exclusion and inclusion in the sample cluster. Geocoding errors of exclusion affect the P-sample nonmatch rate and geocoding errors of inclusion affect the E-sample erroneous enumeration rate. If the geocoding error omits the census housing unit from the sample block cluster, the P-sample people and housing units will not be matched. Conversely, if the geocoding error includes the census housing unit in the sample block cluster, the E-sample people will be erroneously enumerated.

Seventeen clusters were selected for extended search in dress rehearsal. Ten clusters were in South Carolina and seven were in Sacramento. There were none in Menominee. The criteria for selecting targeted extended search clusters was based on the absolute value of the difference in the number of A. C. E. housing units coded as not matched to the census and the number of census housing units classified as geocoding errors. Clusters were identified for extended search when this difference was 100 in Sacramento and 50 in South Carolina. The two clusters in South Carolina with the largest weighted difference were also included because their sampling weight was large (i.e., 2,857.14 and 283.32).

The clusters selected for targeted extended search for the 2000 Accuracy and Coverage Evaluation are:

- Clusters included with certainty
  - Relisted clusters in A. C. E.
  - Five percent of clusters with the most census geocoding errors and A. C. E. address nonmatches
  - Five percent of clusters with the most weighted census geocoding errors and A. C. E. address nonmatches

- Clusters selected at random from the clusters with A. C. E. housing unit nonmatches (i.e., A. C. E. housing units coded CI or UI) and census housing units identified as geocoding errors (i.e., coded GE).

Clusters without A. C. E. housing unit nonmatches and census geocoding errors are out-of-scope for the targeted extended search sampling. The initial housing unit matching results are used to identify the A. C. E housing unit nonmatches and census housing unit geocoding errors. Any changes to the census inventory of housing units is not reflected in the housing unit matching used to identify targeted extended search clusters.

The Puerto Rico clusters will be selected separate from the fifty states and the District of Columbia. List/enumerate clusters will not be included in this part of the cluster selection for targeted extended search. Special procedures for list/enumerate areas will be developed as needed for clusters with high person nonmatch and geocoding error rates.

The search area is expanded to include one ring of blocks surrounding the cluster. A block was in the first ring of blocks if the block touched the cluster of sample blocks at one or more points. This includes the blocks that touch the corner of a sample block.

The search area is expanded for these clusters for the P-sample whole household nonmatches with no address match and for the E-sample people in housing units coded as geocoding errors. This extended search is targeted at the clusters most likely to benefit from expanding the search area. In addition, the work is targeted to blocks within the search area where the geocoding error is located. In 1990, errors were made because matching and duplicate search in the surrounding blocks was not a common event. There were many housing units in surrounding blocks to search for matching people and coding census duplicates. There was anecdotal evidence of clerks who did not bother to look in surrounding blocks because they rarely found anything. We are targeting the expanded searching to eliminate clerical errors. The clerks know there is a good chance of finding matches or duplicates when we expand the search area when there is a possibility of geocoding error and we limit the search in the expanded search area to blocks containing geocoding errors.

The process is outlined below and a flowchart of the process is attached in Attachment 3.

### 4.7.1 P-sample Matching Extended Search

The search area is expanded to search for the P-sample whole household nonmatches with no address match in the surrounding blocks for the targeted extended search clusters[18]. The clerical matching is conducted in one ring of blocks surrounding the sample block cluster. If the basic street address is located in the surrounding blocks, person matching will be conducted in the block where the basic street address is located. The matching is also conducted when there is a possible address match in a surrounding block. Matches to people enumerated in the surrounding blocks will be coded M. Possible matches to people enumerated in the surrounding blocks will be coded P. This process only reduces the P-sample nonmatch rate when there is geocoding error in the census and the housing unit is incorrectly counted in the surrounding ring of blocks.

The matches and possible matches to census people in surrounding blocks can be identified by the E-sample indicator. The matches and possible matches to census people with an E-sample indicator of 3 are in the first ring of blocks surrounding the sample block.[19]

The census people in targeted blocks are available for matching, but no further processing is required if they do not match. Since everyone must have a code, they are assigned a code of N3. If the clerical matcher needs to remove a link, the software automatically assigns the N3 code for the census person in the surrounding block who does not match. In addition, the non E-sample people with insufficient information for matching have been assigned a code of K3. The K3 code is used for both census people in surrounding blocks in housing units and in group quarters. People with the K3 code are suppressed from clerical matching. The K3 code is defined for completeness and will only be on computer data files. The K3 code will not be assigned clerically. There is no need to review the images for the census people in surrounding blocks coded K3, since we do not have images for the surrounding blocks. Any non E-sample person who is identified clerically as having insufficient information for matching will remain coded N3 and will not be in the clerical matching.

[18] There is no searching in surrounding blocks for partial household nonmatches or for whole household nonmatches with matching addresses.

[19] See Section 6.7 for definitions of the E-sample indicators.

The census people enumerated in group quarters in the surrounding blocks are included in the matching. Census people in group quarters in the surrounding blocks and do not match to the P-sample people have an E-sample indicator of 3 and a match code of Q3.

All three codes, N3, K3, and Q3 are computer codes. The clerks will not use these codes. The clerical matchers can only assign a match or possible match to these people in surrounding blocks. They will never see the census people coded K3.

### 4.7.2 E-sample Extended Search for Geocoding Errors

The search area is expanded for the targeted extended search clusters with housing units identified as geocoding errors in the housing unit phase of A. C. E. A field visit is conducted in targeted extended search clusters with geocoding errors to identify the housing units that exist on the ground in the surrounding blocks. These housing units should be counted in a block in the surrounding blocks when found as actually existing in that block. This visit is conducted in the same general time frame as the A. C. E. person interview. If the housing unit exists in the surrounding blocks, the clerks will code the E-sample person as geocoded to the surrounding blocks (i.e., GS) during the before follow-up person matching. If the housing unit exists in the sample block cluster, the E-sample person is coded as not a geocoding error, because that housing unit exists in the sample cluster after all (i.e., GC). A person follow-up interview for the E-sample nonmatches coded GS and GC is needed to identify other reasons for erroneous enumeration, such as fictitious people and other residences. If the housing unit does not exist in the surrounding blocks or can not be located on the map sent with the case, the E-sample person is coded as GE, indicating a person who is erroneously enumerated because the housing unit is incorrectly geocoded in the initial enumeration.

If the field work was not done or if we can not determine if the block number entered on the form is in the block cluster or in the surrounding blocks, the unresolved code, GU is used. There is no follow-up for the unresolved cases.

### 4.7.3 E-sample Targeted Duplicate Search

A limited search for duplicated people is conducted in the targeted extended search clusters with geocoding error in the surrounding blocks. This duplicate search is to identify people duplicated because of geocoding error. This duplicate search is done first on housing units and then on people.

If the housing unit is identified as geocoded in the surrounding blocks (i.e., GS), a housing unit duplicate search is conducted in the block where the GS should have been counted. If the housing unit is duplicated, a search is conducted to identify people duplicated. The duplicate search is conducted in the block where the duplicated housing unit is located. These people are duplicated because the housing unit was enumerated correctly in a surrounding block and incorrectly in the sample block cluster. If the housing unit is not duplicated, a search for person duplication is not conducted. We are only interested in the person duplication in duplicated housing units caused by housing unit geocoding error in the surrounding blocks. In practice, if the basic street address is also in the surrounding blocks, a duplicate search will be conducted. The duplicate search is also conducted when there is a possible address match in a surrounding block. There will be no searching for duplicates in the group quarters enumerations.

The code identifying duplication (i.e., DE) is also used to identify duplication between the E-sample (i.e., E-sample indicator 1) and a census enumeration in surrounding blocks (i.e., E-sample indicator 3). When an E-sample person (i.e., E-sample indicator 1) who was coded as GS is found to be duplicated in the surrounding block, the E-sample person is coded as DE and the person in the surrounding block remains coded N3. The duplicates are linked.[20]

If the follow-up concludes the people are not erroneously enumerated and no duplication is found, these census people are correctly enumerated within the expanded search area. An advantage to this processing of the targeted extended search clusters with geocoding error is the follow-up is done before the person matching begins. All information is given to the matching clerks to code the people in targeted extended search block clusters.

### 4.7.4 Added and Deleted Census Housing Units

A limitation with this process is for census housing units added to the block cluster in clusters selected for targeted extended search since the initial housing unit processing. These housing units have not been included in housing unit matching. Follow-up for the E-sample housing unit has not been conducted. Therefore, the geocoding errors are not known. If we have a cluster that is not identified for targeted extended search and a large building is added to the cluster, the first time they enter the picture is during person matching. We will identify geocoding errors during the person follow-up. If any of these cases should have been included in the targeted extended search and are incorrectly geocoded, we would need to generate another follow-up to identify the ones that actually exist in the surrounding blocks from the ones that exist outside the expanded search area.

There is not sufficient time to conduct another interview to determine which census housing units with geocoding error really exist in the first ring of surrounding blocks. The original plan was to code these people with census geocoding errors as unresolved, but the Missing Data Team concluded there was not a good donor pool to use to impute the probability of correct

[20] The code indicating duplication is assigned in three situations: (1). There is duplication between two E-sample people. One person with E-sample indicator of 1 is assigned the DE. (2). There is duplication between the E-sample person (i.e., E-sample indicator 1) and a census person in the sample block cluster subsampled out and not in the E-sample (i.e., E-sample indicator 2). The person with E-sample indicator 2 is assigned the code DE. (3). There is duplication between an E-sample person and a person in the surrounding blocks (i.e., E-sample indicator 3). The person with E-sample indicator 1 is assigned the DE.

enumeration for these people.

As a consequence, we could not include in the targeted extended search the P-sample people in P-sample housing units that matched to a census housing unit in the original housing unit matching, but the P-sample people are all not matched because the matching census housing unit was deleted from the final census inventory of housing units in the CUF. A balancing bias would have occurred unless we exclude both types of census changes from the targeted extended search.

The result is :

- All E-sample housing unit discovered to have been added to the block cluster with incorrect geography are coded GE. This applies to clusters in the targeted extended search and clusters not in the targeted extended search.
- There will be no surrounding block searching for P-sample people in whole households of P-sample not matched people where the matching census address was deleted before the CUF was created. Therefore, we do not need to check for this type of housing unit. This type of household will not be searched in surrounding blocks because the whole household of not matched people had an address match in the initial housing unit matching and the housing unit match code is matched.

### 4.8 Technician Review

The A. C. E. Technicians review cases coded RV by the clerical matchers. Other clusters with a cluster review indicator, such as high weights and high rates of nonmatch and unresolved are reviewed by the Technicians as a quality assurance activity. In addition, the initial work by the clerical matchers is included in the quality assurance at 100 percent. The amount of work reviewed before the clerical matcher is qualified to continue is not yet determined. It probably should be a number of clusters worked and should include a large cluster.

When the clerical matcher has passed the quality assurance, the later work will be selected for quality assurance at a rate of 10 percent of the clusters completed by the clerical matcher. If the clerical matcher never passes the quality assurance, the work of that clerical matcher will be reviewed at 100 percent. We will try to reassign the clerical matchers that do not qualify to do the clerical matching.

### 4.9 Analyst Review

The A. C. E. Analysts review cases coded RV by the Technicians. The Technician's work will be reviewed for quality assurance in the same manner as the clerical matchers.

## 4.10 A. C. E. Person Follow-up

The person follow-up is conducted to gather additional information to accurately code the residence status of the nonmatched P-sample people and the enumeration status of the E-sample people. The P-sample nonmatches do not match to the census. We want to make sure these P-sample nonmatches actually lived in the sample block cluster on Census Day. The P-sample nonmatch is sent for a follow-up interview when there is a possibility the residence status is not correct, such as partial household nonmatches, whole household nonmatches when the interview was obtained by a proxy interview, and when there is a conflicting household situations (i.e., Smith/Jones cases). The E-sample nonmatches are sent for a follow-up interview to determine if they were correctly or erroneously enumerated in the block cluster. We send possible matches for an interview to resolve their match status. There are other cases sent, such as matched people with unresolved residence status and other types of cases considered to have the potential for geographic errors in the P-sample.

The following cases were sent to person follow-up in Dress Rehearsal:

- P-sample partial household nonmatches
- P-sample whole household nonmatches where the census enumerated different E-sample people (i.e., conflicting households or Smith/Jones cases)
- P-sample whole household nonmatches where the housing unit does not match and the A. C. E. person interview was with a proxy respondent
- P-sample whole household nonmatches where the housing unit does match and the A. C. E. person interview was with a proxy respondent
- E-sample nonmatches[21]
- Possible matches[22]
- P-sample matches and nonmatches with unresolved residence status

In cases where there is a P-sample nonmatch that is not identified for follow-up, but there is a possible match requiring follow-up, the P-sample nonmatch will be sent for a follow-up interview. These people are interviewed for follow-up, since we are already visiting the housing unit for the interview with the possible match. There should only be a few of these cases. In addition, if there is P-sample nonmatch with unresolved residence status in a household with other P-sample nonmatches not planned for follow-up, all P-sample nonmatches will go to follow-up. The same logic applies in this case also.

During the T-night operation, census questionnaires are distributed to places like marinas, carnivals, racetracks, campgrounds, RV parks, etc. If the people in these places have no usual

[21] The GS and GC codes are considered as E-sample nonmatches and are sent for a follow-up interview. The code GU is unresolved and does not need a follow-up interview.

[22] Possible matches to all E-sample indicators are sent for a follow-up interview.

residence, the people are enumerated on enumerator census questionnaires and a housing unit record is created for these people in the DMAF. If they indicate they have a usual home elsewhere, the census questionnaires are processed in the primary selection algorithm and the people are enumerated at their usual home.

The people enumerated in these housing units will be included in the person matching. The housing units look like any other housing unit, except the "Just in Case" box 3 is marked. The census people will not go for a follow-up interview, because these people are highly mobile and there would be a high rate of unresolved. Therefore, if there is a whole household E-sample nonmatch and the housing unit has the "Just in Case" box 3 marked, the code for the people in the household will be changed from NE to UE.

The whole and partial codes and housing unit matching codes are already being generated by computer. The conflicting household code needs to be created. When there are both P-sample whole household nonmatches and E-sample whole household nonmatches in housing units that matched, the P-sample household is conflicting.

Therefore, we will not be sending the whole household P-sample nonmatches that are P-sample interviews with a household member and have no census people (i.e., vacant or whole household imputations). In addition, we will not conduct follow-up when there is no housing unit match to the census and the P-sample interview was conducted with a household member. This will reduce the follow-up of P-sample nonmatches by 53.2 percent according to the Chicago data.

The following table summarizes which P-sample nonmatches will be sent for a follow-up interview:

| Type of P-sample Nonmatch | Person Interview with a Proxy Respondent | Person Interview with a Household Member |
|---|---|---|
| Partial household nonmatch[23] | Followed up | Followed up |
| Whole household nonmatch where the housing unit does not match[24] | Followed up | Not followed up |
| Whole household nonmatch where the housing unit is matched to the census[25] | Followed up | Not followed up |
| Whole household nonmatch with conflicting households[26] | Followed up | Followed up |

These codes are not perfect, but the best that can be generated. There are added housing units and other reasons why the conflicting code is not perfect, but this should not be a large number. We know that doing this coding clerically in 1990 did not work. The consequence of these codes not being perfect is in the analysis of the data and in the imputation. The different types of P-sample nonmatches are tabulated by whole or partial household nonmatch and by the match status of the housing unit. If the E-sample housing unit is added since the initial housing unit matching, a P-sample nonmatch could be misclassified as not matched to the census when the housing units will be matched in the final housing unit matching. The imputation methodology

[23] The P-sample nonmatch is in a household where at least one P-sample person has been matched to a person enumerated in the census.

[24] The address for the P-sample housing unit is not matched to an address in the census. The whole household of P-sample people are not matched to people enumerated in the census, because the housing unit is not matched in the census. The housing unit is not necessarily missed in the census. It is not found within this block cluster.

[25] The address for the P-sample housing unit is matched to an address in the census, but there are no census records for the census housing unit. The census housing unit is either vacant, all people are imputed, or all people are coded as insufficient information for matching and follow-up for A. C. E. All people are imputed in the census when not enough data is collected to create census person records for the whole household. The minimum amount of data collected for a household could be population count for the household. All people in the census household may have been converted to insufficient information for matching and follow-up and coded the same as the households with no census person records.

[26] The address for the P-sample housing unit is matched to an occupied census housing unit, but the people in the P-sample household are different from the people enumerated in the census. In other words, the people in these households are conflicting. These cases have been referred to as Smith/Jones cases in the past.

uses these reason for nonmatch codes to impute the probability of being erroneously enumerated for E-sample people with an unresolved enumeration status. The P-sample imputation of residence probability depends on whether a person is sent to follow-up. Hence, the imputation of residence probability in the P-sample is affected because the decision to send a P-sample person to follow-up is determined by the match status of the housing unit. If the matching has not been done for added census housing units, the match status may not be correct. This should not have much of an effect on either imputation or dual system estimates because of the small volume of census adds.

If the whole households are identified as not needing a follow-up interview and a person in the household has unresolved residence status, the person is sent for a follow-up interview. In other words, the unresolved residence status overrides the decision not to send the person to follow-up.

The E-sample people that are whole household nonmatches in an E-sample housing unit that was added since the housing unit matching phase, need to have their housing unit enumeration status checked during the person follow-up. We need to identify the geocoding errors, since all of the people in the household are erroneously enumerated in the cluster. If a housing unit is incorrectly geocoded, it did not exist as a housing unit in the cluster on census day. The census housing units in list/enumerate clusters are treaded as added census housing units, since they did not get matched in the initial housing unit matching conducted in the Spring of 2000.

In addition for the 2000 A. C. E. we should add the following types of cases to follow-up to collect more information to verify P-sample housing unit geography. A geography section would be added to the P-sample nonmatch section with questions similar to the ones in the E-sample geography section. The geography section will be required for the following in addition to asking the person nonmatch questions:

- All P-sample whole household nonmatches in relisted clusters, clusters in list/enumerate areas, and any other clusters not included in the housing unit matching phase of the A. C. E.
- All P-sample whole household nonmatches in clusters with a high rate of P-sample person nonmatch. High has been determine to be 45 percent.
- P-sample whole household nonmatches where the interviewer for the person interview changed the address for the P-sample housing unit. Information about the accuracy of the P-sample geography would be obtained.
- Any P-sample whole household nonmatch identified by the Analyst as needing follow-up.

The first three bullets describe clusters that potentially could have P-sample geocoding error. Adding a section to the P-sample nonmatch follow-up to identify geocoding errors would allow us to remove these people from the P-sample to correct clusters that would become outliers later on.

The relisted clusters and list/enumerated clusters have had no housing unit matching. Therefore, the P-sample geocoding has not been verified. There were problem clusters in dress rehearsal where the relisted clusters contained geocoding errors. This is our chance to remove clusters with P-sample geocoding errors in relisted and list/enumerate clusters.

One reason for a high rate of P-sample person nonmatch is P-sample geocoding error. There was a problem cluster in dress rehearsal that had P-sample geocoding error. It was listed in error and the listing quality assurance did not identify the P-sample housing units as being incorrectly geocoded. Adding these whole households of P-sample nonmatches to a new section to collect P-sample geographic information, would allow us to remove these people from the P-sample.

The interviewer can change the address of the P-sample housing unit during the person interview. We would like to identify situations where an interviewer is in the wrong place and in trying to fix the cluster really is relisting a cluster during the person interview. Sending these whole household nonmatches to follow-up to collect information about census geography, would allow us to detect these errors and correct them.

We should also allow the analysts to send any case to follow-up in 2000 in case something unusual comes up.

If the follow-up interview identifies a P-sample housing unit as a geocoding error, the people and housing unit will be removed from the P-sample. The whole households of P-sample people incorrectly listed in the cluster will be coded as P-sample geocoding error (i.e. GP).

### 4.10.1 The A. C. E. Person Follow-up Questionnaire

The person follow-up is conducted using a paper questionnaire. This paper questionnaire is prepared using Docuprint. The questionnaire is designed to gather information that may resolve matching and residence status problems. For example, a match between P-sample and E-sample people might be made if another piece of information is known. Also, information may be needed to confirm residence status on census day for matched or unmatched people. During the follow-up interview, interviewers will attempt to gather the information needed to code each person as a matched resident/non-resident or a nonmatched resident/non-resident of the block cluster on census day. More emphasis is being made in this questionnaire to obtain a good respondent before the interview in each section is complete.

The clerical preparation required in the past was error prone because it was a tedious task. It was also time consuming. Docuprint will eliminate these clerical errors and speed up the whole process.

### 4.10.2 A. C. E. Person Follow-up Outcome Codes

**Outcome 1 (Completed Interview):**

- An interview is complete if either of the following definitions of "DONE" are true for ALL persons in a case:

1. If the respondent, answers 'yes' to Question 1- (Do you know ......?) and knows where the person was living on census day (Question 4)

Or

2. If the interviewer has successfully contacted 3 knowledgeable people, but still no one knows the person. (The person will be considered DONE and given an after follow-up code of 'Fictitious'.)

**Outcome 2 (Partial Interview):**

- At least 1 person, NOT ALL, is 'DONE' according to the above definitions. This person can be from any section of the questionnaire- i.e.- an A. C. E. person, census person or possibly matched.

**Outcome 3 (Refusal)**

-Unable to complete interview due to a refusal by the respondent who is an occupant of the sample household and will not allow the interview to begin or will not provide the data needed.

**Outcome 4 (Not at home)**

Unable to complete interview due to an inability to contact a member of the household at that particular time.

**Outcome 5 (Callback appointment)**

Interviewer will return at a later time/date to complete the interview.

**Outcome 6 (Other)**

Unable to complete an interview for some other reason than what is listed in Outcome 1-5.

Assumptions

A. 10 is the maximum number of attempts per case (interviewer must complete section 2a, 2b, or 2c, to qualify as an attempt) to locate knowledgeable respondents.

B. If the respondent is knowledgeable, but does not know the person, then this respondent will qualify as 1 of the 3 knowledgeable respondents needed before the person can be considered 'DONE'. The interviewer will continue to ask about this person until three knowledgeable respondents are found whom do not know the person, or, the maximum number of attempts have been satisfied.

## 4.11 After Follow-up Coding

After the follow-up is completed, the results of the interview are reviewed and codes entered into the system by the matching clerks. The codes assigned to the people sent for a follow-up interview are as follows:

P = There is not enough information collected to determine if the possible match is a match or not. The match status of the P-sample and E-sample people is unresolved.

CE = The E-sample person is identified as correctly enumerated from the A. C. E. person follow-up interview.

EE = The E-sample nonmatch is identified as erroneously enumerated from the A. C. E. person follow-up interview.

FE = The E-sample nonmatch is determined to be fictitious in this block cluster during the follow-up interview. The person may have existed, but could not find anyone in this cluster who knew them after talking to at least three people. The E-sample person is erroneously enumerated in the census in this block cluster.

FP = The P-sample person is fictitious in this block cluster. The person was interviewed in error during the person interview. This person is not included in the final P-sample.

UE = Not enough information is collected during the A. C. E. person follow-up interview to identify the census person as correctly or erroneously enumerated in the census. The enumeration status for the E-sample person is unresolved. This code is also used when the E-sample person is followed up to collect geographic information and that information is not collected.

NR = The P-sample person is identified as a resident in the block cluster on census day during the A. C. E. person follow-up interview. The P-sample person is missed in the E-sample.

NL = The P-sample person did not live at the sample address or in the block cluster on census day and was listed as a nonmover or outmover in error.

This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.

NN = The P-sample person is identified as a nonresident in the block cluster on census day during the A. C. E. person follow-up interview, because the person lived in group quarters on census day or had another residence where the person should have been counted on census day according to census residence rules. This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.

NU = Not enough information is collected during the A. C. E. person follow-up interview to identify the P-sample person as a resident or nonresident in the block cluster. The residence status for the P-sample person is unresolved. This code is also used when the P-sample person is followed up to collect geographic information and that information is not collected.

KP = Match not attempted for the P-sample person, because (1) the name is blank or incomplete, such as "Mr. Jones" or (2) the name is not a valid name, such as "White Female" or "Donald Duck".

KE = Match not attempted for the census person. The name is incomplete or not a valid name, such as "Child Jones", or "Mickey Mouse".

DP = The P-sample person is a duplicate of another P-sample person.

DE = The E-sample person is a duplicate of another E-sample person. The non E-sample person in the block cluster (i.e., E-sample indicator 2) is a duplicate of the E-sample person. The E-sample person is a duplicate of a census person in a surrounding block (i.e., E-sample indicator 3).

GP = The P-sample person is removed because the person interview was conducted at a housing unit that exists outside the sample block cluster. The person follow-up identified this housing unit as a P-sample geocoding error.

MR = The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is a resident. The person is in the P-sample and is matched to a census person. If the census person is in the E-sample, the E-sample person is a correct enumeration.

MN = The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is not a resident in this housing unit or in this block cluster. The person is no longer a P-sample person. If the match was to an E-sample person, the E-sample person is an erroneous enumeration.

MU = The A. C. E. person follow-up interview obtained no useful information to resolve the unresolved residence status for the matched person. The P-sample person's residence status is unresolved. If the match was to an E-sample person, the E-sample person's enumeration status is unresolved.

RV = The match status of this P-sample or E-sample person is not clear. A review by either an A. C. E. Technician or Analyst is needed to resolve this case.

N2 = The non E-sample census person in the large block was a possible match and the person follow-up interview determined the name for the P-sample person and name for the person with E-sample indicator of 2 did not refer to the same person. The non E-sample person is assigned the N2 code. Only census people with E-sample indicator of 2 can be assigned the N2 code.

N3 = The census person in the surrounding block was a possible match and the person follow-up interview determined the name for the P-sample person and name for the census person with E-sample indicator of 3 did not refer to the same person. The census person in the surrounding block is assigned the N3 code. Only census people with E-sample indicator of 3 can be assigned the N3 code.

Q2 = The non E-sample census person in group quarters in the large block was a possible match and the person follow-up interview determined the name for the P-sample person and name for the group quarters person with E-sample indicator of 2 did not refer to the same person. The non E-sample person in group quarters is assigned the Q2 code. Only census people with E-sample indicator of 2 in group quarters can be assigned the Q2 code.

Q3 = The census person in group quarters in the surrounding block was a possible match and the person follow-up interview determined the name for the P-sample person and name for the census person in group quarters with E-sample indicator of 3 did not refer to the same person. The census person in group quarters in the surrounding block is assigned the Q3 code. Only census people in group quarters with E-sample indicator of 3 can be assigned the Q3 code.

The code for possible match (i.e., P) is allowed to E-sample indicators of 1, 2, or 3. The NU, UE, and MU codes are used when the person did not live there on census day, but did not give another address where they did live and we cannot determine if they moved within the cluster or not. The NU, UE, and MU codes are also used when the address given is not complete enough to determine if it is in the block cluster or not, such as Route 1 or Elm Street. The NU or UE is used when the geographic work for the person followed up is not done.

The clerks, technicians, and analysts are allowed to correct before follow-up codes, when errors in the before follow-up matching are discovered during the after follow-up matching and coding. The allowable codes are KP, KE, DP, and DE. These before follow-up codes are entered into the database in the after follow-up stage of the matching. The matchers will not be doing rematching to look for errors, but there will be situations where we followed up a person who should have been insufficient information for matching, for example. Matches and duplicates may also be discovered during the after follow-up coding and can be corrected. We are not actively looking for errors, but they may be corrected when discovered.

The NC code is not used in after follow-up coding. This work should have been done in the before follow-up matching phase. There is no reason to code NC, since the code is only used to reduce the follow-up workload. Images are available in case someone wants to look at them, but the image work should have been done in the before follow-up matching phase.

Duplicate coding is not actually allowed in the batch processing. Duplicate coding is only allowed during the cluster processing by the technicians and analysts for both the P-sample and the E-sample. The clerical matchers will assign an RV code if they discover a duplicate, but this should be rare when only looking at one household at a time.

There are unresolved codes assigned to P-sample people. There are codes to indicate that a P-sample person has unresolved match status and codes to indicate that a P-sample person has unresolved residence status. The probability of being matched is imputed for the P-sample people with unresolved match status. The probability that the P-sample person is a resident is imputed when the follow-up did not give enough information to resolve the person's residence status. The probability that a P-sample person is a resident is the probability that the person is included as a P-sample person.

The N2, Q2, N3, and Q3 codes are used in the after follow-up matching and coding, but will be automatically coded by the software.

The codes used in surrounding block matching are also allowed in the after follow-up coding in list/enumerate clusters selected for targeted extended search. These codes are.

GS = The E-sample person is enumerated in a housing unit that exists in a surrounding block.

GC = The E-sample person is enumerated in a housing unit that exists in the sample cluster.

GE = The E-sample person is erroneously enumerated in this block cluster, because the housing unit is a geocoding error.

GU = The geographic work for the targeted extended search is unresolved. The field work for the targeted extended search was not done or the block number on the form was not in the surrounding blocks, in the block cluster, or on the map. It is not clear where the housing unit is located.

### 4.12 Outlier Review

The outlier review is conducted in two pieces: cluster review and post-strata review. The quality assurance will identify clusters for additional work. A score is calculated for each cluster. All clusters above a cut-off are included in the automated outlier review.

For each cluster we calculate the score as:

$$\text{SCORE=}\ \left(\sum_{PSN} PSW + \sum_{PSU} PSW + \sum_{ESU} ESW + \sum_{EE} ESW\right) \Big/ \sqrt{T}$$

Where:

PSN are the P-sample nonmatches, i.e. those with P-sample match codes NP, NC, NU, NR

PSU are the P-sample unresolveds, i.e. those with P-sample match codes KI, KP and all persons with match codes P and MU

EE are the E-sample erroneous enumerations, i.e. those with E-sample enumeration codes EE, GE, KE, DE, FE, MN

ESU are the E-sample unresolveds, i.e. those with E-sample enumeration codes UE, GU

T is the unweighted sum of those persons with match codes M, MR, MU, NP, NC, NR, NU, P, KI, KP, CE, GE, EE, FE, DE, KE, UE, GU, MN

PSW is the final P-sample cluster weight
ESW is the E-sample cluster weight

The Analysts conduct an outlier review for the clusters with really high weighted nonmatch and erroneous enumeration rates. Journals will be written for these outlier clusters. We need documentation for all of the outlier clusters. Errors in the cluster, such as matching, geocoding, or listing errors, will be corrected and documentation of the error will be written. Clusters without errors will be reviewed and any documentation of the cluster will be written. This review and documentation is much like the outlier cluster review conducted after the Post Enumeration Survey work was completed. In 1991, the outlier cluster review was conducted for 104 clusters out of 5290 or about two percent of the clusters. This outlier review and documentation will be conducted before the matching has been completed and the cluster closed out. This documentation could be used to identify clusters for downweighting.

The second piece to the outlier review is done by post-strata. The data will be examined on a flow basis. Programs will be written at headquarters to calculate the weighted nonmatch and erroneous enumeration rates by post-strata and other variables that we have on the files or derived from data on the files. The correct enumeration rate divided by the match rate is a measure of the net undercount. This will give us advance warning of post-strata with potential problems before the estimation phase has been completed. If we had this in place in dress rehearsal, we would have known about the tenure problem in rural South Carolina several months early. We would have been able to start investigating the problem earlier and have documentation ready before the estimation phase is completed.

We will not restrict this work to the after follow-up person matching. This work can be done in all phases of the matching. We need the before and after follow-up data on a flow basis for housing unit and person matching. We can do things like write duplicate identification programs for the before follow-up matching to identify clusters where they did not do the duplicate search. We will start identifying problem clusters in the housing unit phase. We will look at the before follow-up housing unit and person matching results.

## 4.13 Additional Computer Coding

There are additional codes assigned to each P-sample and E-sample person to be used in the imputation and for analysis purposes. These codes are Whole/Partial Code, Address Code, and Basic Street Address Code. This coding is done by computer after all data are entered for the cluster in the before follow-up matching and again after the follow-up codes have been entered into the database.

### 4.13.1 Whole/Partial Code

| | | |
|---|---|---|
| 1 | = | Partial household nonmatch |
| 2 | = | Whole household nonmatch |
| 3 | = | Whole household match |
| 4 | = | All other |

A code to identify whole household matches, whole household nonmatches, partial household nonmatches, and other is created by computer after the clerical matching is completed. Each household is assigned a code of partial household match when there is at least one nonmatch and at least one match. Each household is assigned a code of whole household nonmatch when all members of the household are nonmatches. Each person is assigned a code of whole household match when all members of the household are matches. All of the codes in the other codes column are ignored when assigning codes 1 through 3 in the whole/partial code. Each household is assigned a code of other when the above three codes do not apply. These codes are assigned for both the P-sample and the E-sample. The whole/partial code is created before follow-up and after follow-up.

**Classification of Match, Nonmatch, and Other Codes**

| | Match | Nonmatch | Other Codes |
|---|---|---|---|
| P-sample | | | |
| Before Follow-up | M | NP, NC | P, KI, KP, DP |
| Final | M, MR, MN, MU | NP, NR, NL, NN, NU, FP, NC | P, KI, KP, DP, GP |
| E-sample | | | |
| Before Follow-up | M | NE, GE, DE, GU, GS, GC, FE | P, KE |
| Final | M, MR, MN, MU | CE, EE, UE, GE, DE, FE, GU | P, KE |

### 4.13.2 Address Code

1 = The housing unit matched during the Housing Unit Phase.
2 = The housing unit is not coded as matched during the Housing Unit Phase.
3 = The housing unit is added to the CUF after the housing unit phase of A. C. E.
4 = Conflicting households.

These codes are assigned to each housing unit to indicate matched, not coded as matched, or added to the census after the housing unit phase of A. C. E. The ones not coded as matched have any code except a code of matched.

A conflicting household is where the A. C. E. housing unit matches the census housing unit and both contain whole households of nonmatched people. There is a conflict because the A. C. E. contains one household and the census contains another household. These cases have been referred to as Smith/Jones households in the past, because the Smith's are in the A. C. E. and the Jones's are in the census at the same address. These contradictions are resolved during the follow-up interview where it will be determined who actually lived at the sample address on census day.

There may be cases where there are nonmatches in the P-sample, nonmatches in the E-sample, and a possible match. The number of cases should be small. The whole/partial code will be whole household nonmatch and the address code will not be conflicting, because there is a possible match in the household.

### 4.13.3 Basic street address Code

1 = Single unit
2 = Multi-unit

The codes indicating single units and multi-units are assigned. The basic street address code for the census people is assigned based on the unit designation in the CUF. If the unit designation is present, the basic street address code is a multi-unit basic street address. If the unit designation is not present, the basic street address code is a single unit structure.

### 4.13.4 Follow-up Flag

blank = No follow-up for the person and no follow-up for the household.
0 = The household is followed up, but the person is not followed up
1 = The person is flagged for a follow-up interview.
2 = The clerks have written a special question for the follow-up interviewer to ask.

## 4.14 Final P-sample Person Match Codes

The probability of being matched is estimated for the P-sample people with unresolved match status.

### 4.14.1 Matched

M = The P-sample and the census people were matched.

MR = The P-sample follow-up interview determined that the matched person with unresolved residence status is a resident. The person is a P-sample person and is matched to a E-sample person.

MU = The A. C. E. person follow-up interview obtained no useful information to resolve the residence status for the matched person who had a residence status of unresolved before follow-up. The P-sample person's residence status is unresolved and the E-sample person's enumeration status is unresolved.

### 4.14.2 Not Matched

NP = The P-sample person is not matched to a census person. There was no follow-up for the whole household nonmatches from person interviews with household members and the whole household nonmatches were not conflicting household nonmatches.

NC = The P-sample nonmatch was found on the census roster. This person in a partial nonmatch household was not matched to the census because only name was collected in the census for this person in a large household and the census person was not data defined. No follow-up interview is necessary.

NR = The P-sample person is identified as a resident in the block cluster on census day during the A. C. E. person follow-up interview. The P-sample person is missed in the census.

NU = Not enough information is collected during the A. C. E. person follow-up interview to identify the P-sample person as a resident or nonresident in the block cluster. The residence status for the P-sample person is unresolved. This code is also used when the a P-sample person is followed up to collect geographic information and that information is not collected.

### 4.14.3 Unresolved

P = There is not enough information collected to determine if the possible match is a match or not. The match status of the P-sample person and the E-sample person is unresolved.

KI = Match not attempted for the P-sample person because the person has insufficient information for matching and follow-up. The name is blank or incomplete or the name is complete but the person has only one characteristic. This is a computer assigned code and these people are suppressed from view by the matchers.

KP = Match not attempted for the P-sample person, because (1) the name is incomplete, such as "Mr. Jones", or (2) the name is not a valid name, such as "White Female" or "Donald Duck". This is a clerically assigned code.

### 4.14.4 Removed from the P-sample

FP = The P-sample person is fictitious in this block cluster. The person was interviewed in error during the person interview. This person is not included in the final P-sample.

NL = The P-sample person did not live at the sample address or in the block cluster on census day and was listed as a nonmover or outmover in error. This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.

NN = The P-sample person is identified as a nonresident in the block cluster on census day during the A. C. E. person follow-up interview, because the person lived in group quarters on census day or had another residence where the person should have been counted on census day according to census residence rules. This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.

DP = The P-sample person is a duplicate of another P-sample person.

MN = The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is not a resident in this housing unit or in this block cluster. The person is no longer in the list of P-sample people and the E-sample person is an erroneous enumeration.

GP = The P-sample person is removed because the person interview was conducted at a housing unit that exists outside the sample block cluster. The person follow-up identified this housing unit as a P-sample geocoding error.

### 4.15 E-sample Person Enumeration Codes

The probability of being correctly enumerated is estimated for the E-sample people with unresolved enumeration status.

#### 4.15.1 Correctly Enumerated

M = The P-sample and E-sample people were matched. The E-sample person is correctly enumerated.

CE = The E-sample nonmatch is identified as correctly enumerated during the A. C. E. person follow-up interview.

MR = The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is a resident. The person is a P-sample person and is matched to an E-sample person.

#### 4.15.2 Erroneously Enumerated[27]

GE = The E-sample person is erroneously enumerated in this block cluster, because the census housing unit is a geocoding error (i.e., counted in the block cluster in error). The E-sample person should have been enumerated elsewhere in the census.

EE = The E-sample nonmatch is identified during the person follow-up interview as erroneously enumerated.

FE = The E-sample nonmatch is determined to be fictitious in this block cluster during the follow-up interview. The person may have existed, but should not have been enumerated in the census within this block cluster. The E-sample person is

[27] The E-sample people who are duplicated with census people with E-sample indicator of 2 are not full erroneous enumerations. If the E-sample person with an E-sample indicator of 1 is duplicated once with a census person with an E-sample indicator of 2, the E-sample person is given one half of an erroneous enumeration. If the E-sample person is duplicated twice with non E-sample people in the cluster, the E-sample person is given two thirds of an erroneous enumeration. The formula is the number of times duplicated is d and the proportion of erroneous enumeration for the E-sample person is d/(d+1). This assumes the E-sample person has been coded as correctly enumerated. If the E-sample person is coded unresolved, the final probability of erroneous enumeration includes an imputation for unresolved enumeration status. If the E-sample person is assigned a match code that indicates erroneous enumeration, the number of times that the E-sample person is duplicated with non E-sample people is irrelevant and ignored. A person can not have a probability of erroneous enumeration that is larger than 100 percent.

erroneously enumerated in the census in this block cluster.

DE = The E-sample person is a duplicate of another E-sample person. The code is also used when the E-sample person is a duplicate of a census person in a surrounding block. The people in the E-sample housing unit are erroneously enumerated because they were counted accurately in the surrounding block and duplicated in the sample block cluster.

MN = The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is not a resident in this housing unit or in this block cluster. The person is no longer in the list of P-sample people and the E-sample person is an erroneous enumeration.

KE = Match not attempted for the E-sample person. The name is blank or incomplete or the name is complete but the person has only one characteristic, which is assigned by computer. The name is incomplete or not a valid name, such as "Child Jones", or "Mickey Mouse", which is assigned clerically.[28]

### 4.15.3 Unresolved

UE = Not enough information is collected during the A. C. E. person follow-up interview to identify the E-sample person as correctly or erroneously enumerated in the E-sample. The enumeration status for the E-sample person is unresolved.[29] This code is also used when the a P-sample person is followed up to collect geographic information and that information is not collected.

MU = The A. C. E. person follow-up interview obtained no useful information to resolve the unresolved residence status for the matched person. The P-sample person's residence status is unresolved and the E-sample person's enumeration status is unresolved.

P = There is not enough information collected to determine if the possible match is a match or not. The status of the P-sample person and the E-sample person is unresolved.

GU = The geographic work for the targeted extended search is unresolved. The code has the same definition in both the before and after follow-up matching. The difference is in after follow-up, the code is only used in the list/enumerate clusters. The field work for the targeted extended search was not done or the block number on the form was not in the surrounding blocks, in the block cluster,

[28] There are two types of insufficient information. The insufficient information for matching and follow-up are coded as KE by the computer or clerically and are treated as erroneous enumerations. The insufficient information in the dual system estimator are whole person imputations and are subtracted from the census count of persons.

[29] The UE code is also used when the person did not live at the sample address on Census Day and the Census Day address is not complete enough to determine if the census day address is in the sample block cluster.

or on the map. It is not clear where the housing unit is located.

### 4.16 Final P-sample Person Residence Status Codes

The probability of being a resident of the housing unit on census day is estimated for P-sample people with unresolved residence status.

#### 4.16.1 Resident

M = The P-sample and the E-sample people were matched.

MR = The P-sample follow-up interview determined that the matched person with unresolved residence status is a resident. The person is a P-sample person and is matched to a E-sample person.

NR = The P-sample person is identified as a resident in the block cluster on census day during the A. C. E. person follow-up interview. The P-sample person is missed in the census.

NP = The P-sample person is not matched to an E-sample person. There was no follow-up for the whole household nonmatches from person interviews with household members and the whole household nonmatches were not conflicting household nonmatches. These people are considered residents of the housing unit on census day.

NC = The P-sample nonmatch was found on the census roster. This person in a partial nonmatch household was not matched to the census because only name was collected in the census for this person in a large household and the census person was not data defined. No follow-up interview is necessary.

#### 4.16.2 Nonresident

FP = The P-sample person is fictitious in this block cluster. The person was interviewed in error during the person interview. This person is not included in the final P-sample.

NL = The P-sample person did not live at the sample address or in the block cluster on census day and was listed as a nonmover or outmover in error. This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.

NN = The P-sample person is identified as a nonresident in the block cluster on census day during the A. C. E. person follow-up interview, because the person lived in group quarters on census day or had another residence where the person should have been counted on census day according to census residence rules. This person is removed from the list of P-sample people, since he or she was collected during the person interview in error.

DP = The P-sample person is a duplicate of another P-sample person.

MN = The A. C. E. person follow-up interview determined that the matched person with unresolved residence status is not a resident in this housing unit or in this block cluster. The person is no longer in the list of P-sample people and the E-sample person is an erroneous enumeration.

GP = The P-sample person is removed because the person interview was conducted at a housing unit that exists outside the sample block cluster. The person follow-up identified this housing unit as a P-sample geocoding error.

### 4.16.3 Unresolved

MU = The A. C. E. person follow-up interview obtained no useful information to resolve the residence status for the matched person who had a residence status of unresolved before follow-up. The P-sample person's residence status is unresolved and the E-sample person's enumeration status is unresolved.

NU = Not enough information is collected during the A. C. E. person follow-up interview to identify the P-sample person as a resident or nonresident in the block cluster. The residence status for the P-sample person is unresolved.[30] This code is also used when the P-sample person is followed up to collect geographic information and that information is not collected.

P = There is not enough information collected to determine if the possible match is a match or not. The match status of the P-sample person and the E-sample person is unresolved.

KI = Match not attempted for the P-sample person because the person has insufficient information for matching and follow-up. The name is blank or incomplete or the name is complete but the person has only one characteristic. This is a computer assigned code only.

KP = Match not attempted for the P-sample person, because (1) the name is incomplete, such as "Mr. Jones", or (2) the name is not a valid name, such as "White Female" or "Donald Duck". This is a clerically assigned code.

### 4.17 Estimation Outcome Codes

Two sets of outcome codes are needed for Procedure C: one for the census day household and one for the interview day household. The final P-sample estimation outcome code identifies the status of the interview for estimation on census day and on the day of the interview. We could have complete interview for the current residents, but a noninterview or vacant for the census day residents.

[30] The NU code is also used when the person did not live at the sample address on Census Day and the Census Day address is not complete enough to determine if the census day address is in the sample block cluster.

### 4.17.1 Final P-sample Estimation Outcome Codes for Census Day

Final P-sample estimation outcome codes are assigned after the follow-up codes are entered into the database. The changes to the estimation outcome codes result from changes to the status of the household identified during the follow-up interview. These codes are used in the imputation and estimation phase of the A. C. E. The final P-sample estimation outcome codes for census day are defined the same way as the preliminary outcome codes, except the final coding is done only for people who have final P-sample match codes of M, MR, MU, NR, NU, NC, NP, P, KI, or KP. The estimation outcome codes are created again without the people who were removed from the P-sample. Housing unit with preliminary estimation outcome codes of 6, 9, 11, or 12 have the same final estimation outcome code. Households with a preliminary outcome code of 10 have the final outcome code of 10. The outcome code may change when people are removed from the P-sample for preliminary estimation outcome codes of 1, 2, or 3. The types of cases where the final outcome codes are changed are:

- Whole households coded NN are converted to vacant (i.e., outcome code 10).
- Whole households coded FP or DP are converted to noninterview (i.e., outcome code 4).
- Whole households coded NL and MN are converted to noninterview (i.e., outcome code 4).
- Households with at least one person coded GP are converted to nonexistent (i.e., outcome code 12).
- Households with a mixture of NN, NL, FP, DP and MN are converted to noninterview (i.e., outcome code 4)
- Households with a mixture of KP, KI, NN, NL, FP, DP, MN are treated as whole households with insufficient information for matching and follow-up (i.e., outcome code 9).

The final estimation outcome codes for Census Day are:

| | | |
|---|---|---|
| 1 | = | Complete interview with a household respondent |
| 2 | = | Complete interview with a proxy respondent |
| 3 | = | Partial interview (i.e., some, but not all P-sample people have sufficient information for matching and follow-up) |
| 4 | = | No Census Day Residents - Households converted to noninterviews |
| 6 | = | Field noninterview |
| 9 | = | No people have sufficient information for matching and follow-up |
| 10 | = | No census day residents - Vacant |
| 11 | = | Vacant on census day |
| 12 | = | Not a housing unit on census day |

P-sample estimation outcome codes 1 through 3 are interviews, 4, 6, and 9 are noninterviews, 10 and 11 are vacant, and 12 is not a housing unit within the block cluster.

### 4.17.2 P-sample Interview Status Codes for Census Day

The P-sample interview status codes for census day are defined as follows:

| | | |
|---|---|---|
| 1 | = | Interview |
| 2 | = | Noninterview |
| 3 | = | Not an occupied housing unit on census day |

If the final estimation outcome code for census day is 1, 2, or 3, the P-sample interview status code for census day is 1. If the final estimation outcome code for census day is 4, 6, or 9, the P-sample interview status code for census day is 2. If the final estimation outcome code for census day is 10, 11, or 12, the P-sample interview status code for census day is 3.

### 4.17.3 Final P-sample Estimation Outcome Codes for Interview Day

The final P-sample estimation outcome codes for interview day are also used in the imputation and estimation phases of A. C. E. The final P-sample estimation outcome codes for interview day are defined the same way as the final P-sample estimation outcome codes for census day, except the final coding is done on the final P-sample people who are residents of the housing unit on interview day. The following codes will define the residents of the housing unit on interview day:

- Inmovers who were not born since census day - the A. C. E. mover status code is 2 and the A. C. E. born since census day code is 1, which is the same as A. C. E. status equal to I = Inmovers.

- Nonmovers who were not born since census day - the A. C. E. mover status code is 1 = nonmover and the A. C. E. born since census day code is 1 = not born since census day.

The final P-sample estimation outcome codes for interview day are:

| | | |
|---|---|---|
| 1 | = | Complete interview with a household respondent |
| 2 | = | Complete interview with a proxy respondent |
| 3 | = | Partial interview (i.e., some, but not all P-sample people sufficient information for matching and follow-up) |
| 5 | = | Refusal |
| 7 | = | Unable to contact a knowledgeable respondent |
| 8 | = | Language problems |
| 9 | = | No people have sufficient information for matching and follow-up |
| 10 | = | No interview day residents (All people are nonresidents.) |

| | | |
|---|---|---|
| 11 | = | Vacant on interview day |
| 12 | = | Not a housing unit on interview day |

P-sample estimation outcome codes 1 through 3 are interviews, 5, 7, 8, 9, and 10 are noninterviews, 11 is vacant on interview day, and 12 is not a housing unit within the block cluster in interview day.

### 4.17.4 P-sample Interview Status Codes for Interview Day

The P-sample interview status codes for interview day are defined as follows:

| | | |
|---|---|---|
| 1 | = | Interview |
| 2 | = | Noninterview |
| 3 | = | Not an occupied housing unit on interview day |

If the final estimation outcome code for interview day is 1, 2, or 3, the P-sample interview status code for interview day is 1. If the final estimation outcome code for interview day is 5, 7, 8, 9, or 10, the P-sample interview status code for interview day is 2. If the final estimation outcome code for interview day is 11 or 12, the P-sample interview status code for interview day is 3.

## 5.0 Final Housing Unit Match

The initial housing unit matching described in Section 3 is conducted before the inventories of census housing units and the A. C. E housing units are final. To produce housing unit coverage estimates it will be necessary to perform a Final Housing Unit Match that incorporates the final state of the E-sample and P-sample housing units. The computer and clerical work for the final housing unit match is currently scheduled to be done after the person matching is completed. The results of the final housing unit match are used to calculate housing unit coverage estimates, to weight the long form estimates and for evaluation efforts.

The Housing Unit Match is to be conducted between the A. C. E. address listing and the January 2000 Census Housing Unit inventory. Only the updates to the Census and A. C. E. housing unit inventories that arise between the time of the Housing Unit Match and the definition of the P-sample and E-sample will be processed during Final Housing Unit Match. In addition, housing unit matching will be conducted for the first time in relisted clusters and list/enumerate clusters during Final Housing Unit Match.

There is no computer matching for the final housing unit matching. However, there is a computer processing stage that determines what goes to the clerical stage. After the clerical matching, a follow-up interview will be conducted for selected cases that were not followed up during the initial housing unit matching.

Note that with the current A. C. E. schedule, we may not have many of the updates from the update/leave areas included in the DMAF when the initial housing unit matching is conducted. Therefore, the workload for the final housing unit match in update/leave areas may be larger than we expect in other areas.

### 5.1 The Computer Processing

The computer processing stage of Final Housing Unit processes the updates that occur between the Housing Unit Match and the creation of the CUF, and the updates due to the large block subsampling, the E-sample identification, and the person interviewing. First, the census records used for Housing Unit Match are extracted from a January 2000 version of the DMAF extract. On the other hand, the Final Housing Unit Match must use the census records that appear on the CUF, which is the final census inventory of housing units. (If there is late census data the CUF shall refer to the version updated with the late census data. An updated CUF extract would be required.)

There will be some records on the DMAF which will not be on the CUF, which we refer to as deletes. There will also be records on the CUF which will not be on the DMAF, which we refer to as adds. The computer processing identifies these adds and deletes for clerical review. It also unlinks duplicates whose primary are deleted. If an add in the E-sample had people who were coded geocoding error, GE, then the housing unit is assigned the code GE.

The computer processing will also clean-up some P-sample addresses based on the person interviews. When the person interview was not conducted because the P-sample housing unit did not exist at the time of the interview, that unit is removed from the final housing unit matching. P-sample housing units with unresolved housing unit match status can also be converted to resolved based on whether or not an interview was conducted. If the final estimation code for census day indicates an interview was conducted (i.e., 1, 2, 3, or 11) for the unresolved cases, the housing unit match code is CI or M (note: pending a future decision a final estimation code for census day of 10 may also suffice). We are assuming that if the housing unit is occupied or vacant on census day, it existed as a housing unit on census day. If the estimation outcome indicates the address was not a housing unit on census day (i.e., equal to 12), the P-sample housing unit match code is ZI. Housing units with ZI codes are not further processed in final housing unit matching. Also, any updated address information will be added to address information contained in the initial housing unit match file. The results of this cleanup only result in clerical processing if removing a P-sample unit unlinks a match to a census unit.

Lastly, in clusters with large numbers of A. C. E. units, the A. C. E. units, and usually the census units too, are subsampled. The subsampling is followed by the E-sample identification. A. C. E. units that are not in the P-sample are removed from further processing and the Census units that are not in the E-sample are assigned an E-sample indicator of 2 (those in the E-sample are assigned an E-sample indicator of 1). Because of the near perfect overlap of the P-sample and E-sample only very few matches will be unlinked and require being sent to the clerical stage.

## 5.2 Final Housing Unit Match and Follow-up

The final housing unit matching will be done by processing the added and deleted census housing units instead of rematching everything. The highest ranking code[31] from the initial housing unit match is loaded into the database.

The following cases will be flagged for clerical review:

- New Census nonmatches, both E-sample and non E-sample.
- New P-sample nonmatches.
- Duplicate links between E-sample housing units coded DE and non E-sample housing units in all clusters.
- P-sample units nonmatched in initial Housing Unit match in targeted extended search clusters.
- E-sample units coded GE in targeted extended search clusters.

The units in list/enumerate and relist clusters are all treated like new nonmatches. Additional matches and possible matches are identified by the clerical matchers, with matches coded M and possible matches P. A duplicate search is conducted to identify census duplicates created by the added census housing units. If an E-sample unit coded DE is duplicated to an non E-sample unit, the clerk will code the non E-sample unit DE and the E-sample unit an appropriate code other than DE, such as M or CE.

These changes to the A. C. E. and census generate some follow-up, because some of these housing units have never had a housing unit follow-up interview. For example, a housing unit added to the census since the initial housing unit match may not be matched to an A. C. E. unit. This new census housing unit needs a follow-up interview to determine if it was a housing unit in the cluster on census day. The census add could be a geocoding error. When the census add is matched to an A. C. E. nonmatch, no follow-up is necessary. A census delete may have been a match and follow-up is needed for the new A. C. E. nonmatch.

New E-sample nonmatches and new possible matches, and some P-sample nonmatches that were not sent to follow-up in the initial housing unit matching are sent for a follow-up interview for the final housing unit phase of the A. C. E. (Most new P-sample nonmatches will not require followup because they can be coded based on the final estimation outcome code.) These cases are coded NE, P or NI during the final housing unit matching. The follow-up interview will be conducted using the same follow-up form and procedures used in the initial housing unit follow-up, except the reference to census day is included in the follow-up questions. The follow-up

[31] The highest ranking code from the MaRCS software is the last code given to a housing unit. If the computer matched two housing units, no codes are assigned in the later stages of matching. In other words, the codes assigned at all levels of matching are combined into one final highest ranking code.

form in the original housing unit followup does not have any reference to census day because the follow-up interviewing begins before census day.

An after follow-up clerical phase will record the information obtained during the follow-up interview. The same codes and procedures will be used for coding the results of the final housing unit follow-up interviews as was used in the initial housing unit match.

## 5.3 Targeted Extended Search

The targeted extended search in surrounding blocks is conducted for housing units in those clusters identified as selected for targeted extended search in the person matching phase of the A. C. E. The targeted extended search will reduce the variance of the housing unit coverage estimates and will make the housing unit coverage geographically congruent with the person. coverage.

The follow-up forms for the targeted extended search are saved from the person phase. For those units added to the CUF that are determined in final housing unit follow-up to be geocoding errors, perform the follow-up immediately after the Final Housing Unit follow-up is completed. This follow-up for targeted extended search is identical to the operation performed during the person interviewing phase which is described earlier in Section 4.7.2.

All E-sample nonmatches coded GE in targeted extended search clusters are geocoded using information on the Targeted Extended Search Field Follow-up forms. They are coded GE if the unit cannot be located, does not exist or is outside the surrounding block search area (one ring of blocks around the A. C. E. block cluster). GE indicates an erroneous enumeration. If the unit is determined to be in the surrounding block code it GS. If the unit is located in the A. C. E. block cluster code it GC. Both GS and GC are correct enumerations. If unit is coded GS perform a duplicate search in that surrounding block in which the GS unit should have been geocoded to.

For all P-sample nonmatches in targeted extended search clusters search for matches in the surrounding blocks. If matched the P-sample unit is coded M. The matched census unit has an E-sample indicator of 3.

## 5.4 Final Housing Unit Match Codes after Computer Processing

After the computer processing there will be the following codes:

NI = The P-sample address is a new nonmatch to a census address.
NE = The census address in the E-sample is a new nonmatch to an A. C. E. address.
N2 = The census address not in the E-sample is a new non-match.
RV = The census address is either a primary not in the E-sample or a duplicate whose primary not in the E-sample

In addition, the housing units with the following codes from initial housing unit match remain on the files: M, CI, CE, EE, GE, DE, MU, UI and UE. See Section 2.7.1 for definitions of these codes. Note that housing units with the following codes from the initial housing unit match are removed from final housing unit match: GI, DI, ZI, ZM and ZE.

## 5.5 Before Follow-up Match Codes

The codes resulting from the BFU processing of the new nonmatches in computer processing, NI, NE and N2, are M, P, NI, NE, DI, DE and MU. In TES clusters there are also the codes GS, GC, GU, and GE. For units not processed in final housing unit match, the match codes from original housing unit match still apply.

## 5.6 Final P-sample Housing Unit Match Codes

### 5.6.1 Matched

M = The P-sample and census addresses are matches.

### 5.6.2 Not Matched

CI = The P-sample housing unit existed as a housing unit at the time of the follow-up interview and is correctly geocoded in the block cluster. The housing unit is not found in the census.

### 5.6.3 Unresolved Housing Unit Status

MU = The P-sample and census addresses match and there is not enough information on the follow-up form to confirm this match as a housing unit with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview. The match status is matched, but the housing unit status is unresolved.

UI = Not enough information on the follow-up form to assign a code to the nonmatched P-sample housing unit with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview. The match status is not matched, but the housing unit status is unresolved.

### 5.6.4 Removed from the P-sample

ZM = The map spot number associated with a housing unit is in error. A delete code of ZM is entered for that housing unit. This code removes the housing unit from the P-sample.

DI = The housing unit should not have been listed in the P-sample. This address is a duplicate of another P-sample address. This address is removed from the P-sample.

ZI = The P-sample address is incorrectly included in the A. C. E. list of housing units. This error is identified clerically after the A. C. E. list is created and does not need to be sent to the field for an interview. This code removed the address from further processing. The code is also used when the P-sample address did not refer to a housing unit at the time of the follow-up interview. For example, the housing unit burned or the mobile home moved. Another example, the address is commercial property or a special place.

GI = The P-sample housing unit existed as a housing unit at the time of the follow-up interview, but is incorrectly listed in the block cluster. The housing unit is an A. C. E. geocoding error.

### 5.6.5 Unresolved Match Status

P = The P-sample housing unit is a possible match to a census housing unit, but the follow-up interview was inconclusive or incomplete.

## 5.7 Final E-sample Housing Unit Enumeration Codes

### 5.7.1 Correctly Enumerated

M = The P-sample and E-sample housing units are matches.

CE = The E-sample housing unit existed as a housing unit at the time of the follow-up interview and is correctly geocoded in the block cluster. The housing unit is not matched to an A. C. E. unit.

GS = The E-sample housing unit was found to be in the surrounding blocks during the targeted extended search field follow-up. The E-sample housing unit was counted once and only once in the expanded search area and is correctly enumerated within the expanded search area.

GC = The E-sample housing unit was found in the sample block cluster during the targeted extended search field follow-up. It is correctly enumerated in the block cluster.

### 5.7.2 Erroneously Enumerated[32]

EE = The E-sample housing unit is erroneously listed on the CUF, because the address is not a housing unit on census day in the block cluster. For example, the housing unit burned or the mobile home moved.. Another example, the address is commercial property or a special place. Also, the address is nonexistent within the sample block cluster.

DE = The census housing unit is erroneously enumerated in the census. The reason for the erroneous enumeration is the address is duplicated in the census.

GE = The E-sample housing unit existed as a housing unit at the time of the follow-up interview, but is incorrectly geocoded to this block cluster. This housing unit is erroneously enumerated in this block cluster, because of a geocoding error.

### 5.7.3 Unresolved

MU = The A. C. E. and census addresses match but there is not enough information on the follow-up form to confirm this match as a housing unit with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.

UE = Not enough information on the follow-up form to assign a code to the E-sample nonmatched housing unit with certainty. The follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.

P = The E-sample housing unit is a possible match to the P-sample housing unit, but the follow-up interview was inconclusive or incomplete.

GU = The E-sample housing unit is not located in the A. C. E. cluster, but there is not enough information to determine whether it was in the surrounding

[32] The E-sample housing units that are duplicated with census housing units with E-sample indicator of 2 are not full erroneous enumerations. If the E-sample housing unit with an E-sample indicator of 1 is duplicated once with a census housing unit with an E-sample indicator of 2, the E-sample housing unit is given one half of an erroneous enumeration. If the E-sample housing unit is duplicated twice with a non E-sample housing unit in the cluster, the E-sample housing unit is given two thirds of an erroneous enumeration. The formula is the number of times duplicated is d and the proportion of erroneous enumeration for the E-sample housing unit is d/(d+1). This assumes the E-sample housing unit has been coded as correctly enumerated. If the E-sample housing unit is coded unresolved, the final probability of erroneous enumeration includes an imputation for unresolved enumeration status. If the E-sample housing unit is assigned a match code that indicates erroneous enumeration, the number of times that the E-sample housing unit is duplicated with non E-sample housing unit is irrelevant and ignored. A housing unit can not have a probability of erroneous enumeration that is larger than 100 percent.

block. The targeted extended search field follow-up interview was not done, was incomplete, was never sent, had contradictory information, or was a noninterview.

## 6.0 Definitions

### 6.1 Procedure C

The person interview for A. C. E. identifies people who lived in the sample housing unit at the time of the interview and on census day (i.e., nonmovers), people who live in the sample housing unit at the time of the interview, but not on census day (i.e., inmovers), and people who lived in the sample housing unit on census day, but not at the time of the interview (i.e., outmovers).

We used Procedure B in the 1990 PES. Procedure B identified the current residents in each sample housing unit. The census day residence of each person was established. The people who live at the sample address now and on census day were nonmovers and the people who lived elsewhere on census day were inmovers. The census day address for the inmovers was collected, along with other information to identify the census geography for the mover addresses. Procedure B compared the nonmovers to the census enumerations within block cluster and in surrounding blocks. The inmovers were matched at their census day address. This involved the cumbersome operation of identifying the census geography for the census day address and matching to the census at the mover address and in the search area. The number of movers was estimated from the inmovers and the match rate for movers was estimated by matching the inmovers at their census day address.

The match rate for outmovers is a better estimate of the match rate for movers than the match rate for inmovers. We know the geography for the residence of the outmovers and can more easily compare the outmovers to the census enumerations. For Procedure B, we match the census enumerations at the inmover's census day address. The census day address is geocoded to obtain census day geography. Some addresses are not geocodeable. Maps were used in 1990 to geocode the ungeocodeable addresses. In 1990, the processing offices did not even have all of the census maps. Also, the volume of maps is large. In 1990, we asked for descriptions of the location of the census day address, intersecting streets, and the names of the nearest neighbors to aid in the clerical geocoding. The clerical geocoding was time consuming and not always successful. We do not always collect accurate and complete census day addresses.

We used Procedure A in the 1995 and 1996 A. C. E. Procedure A attempts to reconstruct the census day household. The census day household is composed of people who have not moved since census day (i.e., nonmovers) and people who have moved since census day (i.e., outmovers). Procedure A compares the nonmovers and outmovers to the census enumerations within block cluster. The search area was the block cluster in 1995 and 1996. The number of movers is estimated from the outmovers and the match rate for movers is estimated by matching the outmovers.

The number of inmovers is a better estimate of the number of movers than the outmovers, because an interviewed household member has better knowledge of their household than the proxy who is now living there. It is difficult to determine who lived at the sample address for outmovers. The present occupants do not always know who lived at the address before they moved in. The present occupants are not always knowledgeable about the date the previous tenants moved out of the housing unit. We must rely on neighbors to provide the names, demographic data, and responses to census residence questions. The rate of noninterviews for Procedure A will likely be higher than for Procedure B. Therefore, the noninterview adjustment must estimate the number of outmovers and the demographic data for these outmovers when the interview is not successful.

Procedure C uses the best features of Procedures A and B. The nonmovers and the outmovers are matched to the census enumerations within the sample block cluster. The match rate for movers is estimated by matching the outmovers. The number of movers and their characteristics are estimated from the inmovers.

## 6.2 T-Night Enumerations

During the T-night operation, census questionnaires are distributed to places like marinas, carnivals, racetracks, campgrounds, RV parks, etc. If the people in these places have no usual residence, the people are enumerated on enumerator census questionnaires and a housing unit record is created for these people in the DMAF. If they indicate they have a usual home elsewhere, the census questionnaires are processed in the primary selection algorithm and the people are enumerated at their usual home.

The people enumerated in these housing units will be included in the person matching. The housing units look like any other housing unit, except the "Just in Case" box 3 is marked. The census people will not go for a follow-up interview, because these people are highly mobile and there would be a high rate of unresolved. Therefore, if there is a whole household E-sample nonmatch and the housing unit has the "Just in Case" box 3 marked, the code for the people in the household will be changed from NE to UE.

## 6.3 List/Enumerate Areas

The listings of census housing units in list/enumerate areas will not be available in time for them to be processed in the initial housing unit matching phase of A. C. E. 2000. In order to keep the A. C. E. on the same interview schedule as the urban and update/leave areas of the country in Census 2000, there will be no initial housing unit matching in list/enumerated areas. The enhanced list of addresses for person interviewing, which is the P-sample, will be the A. C. E. housing units listed in the A. C. E. listing books. Certain classes of structures may be excluded, such as unfit for habitation, but we will include all types of structures listed in the address listing for A. C. E. in the P-sample for interviewing. There could be a change in the status of the unit between the time of listing and census day.

The A. C. E. listing will start in September 1999 for the 2000 A. C. E. We are interested in the census day status of the address. Therefore, we may need to conduct an A. C. E. person interview at each address in the A. C. E. listing book. The interviewers must be trained to establish the status of the housing unit on census day before conducting the A. C. E. interview.

To obtain housing unit coverage all housing unit matching in list/enumerated areas will be conducted during the final housing unit matching. The final housing unit match follow-up workload in list/enumerated areas will be larger than in urban and update/leave areas where we are processing only the changes to the census housing unit inventory. The A. C. E. sample should be small in 2000 in list/enumerate areas, since the number of housing units in list/enumerate areas is less than 500,000 nationally. There may be as many as 250 clusters in sample for interviewing for list/enumerate areas.

The targeted extended search for list/enumerate clusters will be done differently than the other types of enumeration areas. After the before follow-up matching is conducted, clusters will be selected for surrounding block searching. Headquarters personnel will do the field work for the few clusters identified with high rates of whole household E-sample nonmatches. The matching in clusters with high rates of P-sample nonmatches will be conducted only for the clusters in the targeted extended search. A flag will identify which clusters are selected for surrounding block matching to keep the clerical matchers from matching in surrounding blocks for clusters not selected. The census people in surrounding blocks will be in the software for all list/enumerate clusters, but only used for the small number of targeted clusters.

### 6.4 The Sample

The listing sample was selected to make estimates by states. The anticipated listing sample is below. These numbers are estimates from the existing census housing unit inventory. The actual number of listing from the listing books will replace these numbers.

| Listing Sample for the 50 States and the District of Columbia | | | |
|---|---|---|---|
| | Clusters | Housing Units | Housing Units per Cluster |
| Small | 5,000 | 2,400 | 0.5 |
| Medium | 15,393 | 438,600 | 28.5 |
| Large | 8,388 | 1,539,800 | 183.6 |
| American Indian Reservation | 355 | 8,620 | 24.3 |
| Total | 29,136 | 1,989,420 | 68.3 |

The listing sample for Puerto Rico is as follows:

| Listing Sample for Puerto Rico | | | |
|---|---|---|---|
| | Clusters | Housing Units | Housing Units per Cluster |
| Small | 96 | 30 | 0.3 |
| Medium | 244 | 7,100 | 29.1 |
| Large | 219 | 43,400 | 198.2 |
| Total | 559 | 50,530 | 90.4 |

The sample is selected for small, medium, and large block clusters. Small clusters are defined to contain zero to two housing units, medium clusters contain from 3 to 79 housing units, and large clusters contain 80 or more housing units.

There will be a sample reduction to reduce the number of interviewed housing units to approximately 300,000. Block clusters in the small block cluster sample are listed. After the listing is completed, these small block clusters are subsampled. See Section 2.2.3 for more on small block subsampling. The following table lists the sample after sample reduction and small block subsampling. This is the sample for housing unit matching. Numbers will be inserted in these tables after the sample reduction is completed and a revision to this memo will be issued.

| Housing Unit Sample 50 States and the District of Columbia | | | |
|---|---|---|---|
| | Clusters | Housing Units | Housing Units per Cluster |
| Small with 0 to 9 units | | | |
| Medium with 10 or more units | | | |
| Large | | | |
| American Indian Reservation | | | |
| Total | | | |

| Housing Unit Sample for Puerto Rico | | | |
|---|---|---|---|
| | Clusters | Housing Units | Housing Units per Cluster |
| Small | | | |
| Medium | | | |
| Large | | | |
| Total | | | |

Block clusters will be subsampled for A. C. E. in the large block subsampling when the block cluster contains 80 or more housing units. After the housing unit phase of the A. C. E. is completed, the segments of housing units are formed in large blocks. The segments will be selected and the housing units in these segments will be in the A. C. E. sample for interviewing. See section 2.9 for more details about large block subsampling. Only P-sample housing units are interviewed. The following table contains the P-sample housing units after subsampling. Numbers will be inserted in this table after the large block subsampling is completed and a revision to this memo will be issued.

| **Person Matching Sample for Dress Rehearsal** | | | |
|---|---|---|---|
| | Clusters | Housing Units | Housing Units per Cluster |
| United States | | | |
| Puerto Rico | | | |
| Total | | | |

## 6.5 Dual System Estimation

The dual system estimator is

$$DSE = (C - II)\left(\frac{CE}{N_e}\right)\left(\frac{N_p}{M}\right)$$

or

$$DSE = \frac{(C - II)\left(\frac{CE}{N_e}\right)}{\frac{M}{N_p}}$$

where
DSE = the dual system estimate of the population in housing units
C = the count of people in housing units in the census
II = the count of the people in housing units who are whole person imputations, (i.e. people who are not data defined. See Section 4.4.1)
CE = the weighted estimate of the number of correct enumerations in the E-sample
$N_e$ = the weighted number of E-sample people
M = the estimated number of P-sample matches
$N_p$ = the estimated number of P-sample people

Next, the denominator can be expressed in terms of nonmovers and outmovers for Procedure C. This is the number of matches from nonmovers plus the match rate of outmovers times the number of inmovers over the number of nonmovers plus the number of inmovers. The match rate for the movers comes from the outmovers and the number of movers comes from the inmovers.

$$\frac{M}{N_p} = \frac{M_n + (\frac{M_o}{N_o}) * N_i}{N_n + N_i}$$

where

$M_n$ = the weighted estimate of the number of P-sample nonmovers who matched
$M_o$ = the weighted estimate of the number of P-sample outmovers who matched
$N_o$ = the weighted estimate of the number of P-sample outmovers
$N_i$ = the weighted estimate of the number of P-sample inmovers
$N_n$ = the weighted estimate of the number of P-sample nonmovers

Estimates of the population are made within post-strata. The sum of the estimate of the population within each post-strata is the dual system estimate of the population. The net undercount is the difference in the dual system estimate of the population and the number of people counted in the census. The percent net undercount is the net undercount divided by the number of people counted in the census.

## 6.6 Balancing

The search area is the distance in geographic area for searching in the census for a match to the P-sample people and for duplicates and erroneous enumerations in the E-sample. This distance for searching is determined by the Accuracy and Coverage Evaluation design before matching begins. The size of this distance is theoretically unimportant in terms of expected value of the net undercount. The only requirement is to make the P-sample and E-sample search area consistent for them to balance. The variance of the Dual System Estimates decrease as the size of the search area increases.

The P-sample people found to be enumerated within the search area are considered correctly enumerated in the census. The P-sample people not found within the search area are denoted as missed in the census. The E-sample people are correctly enumerated in the census when they are counted only once within the search area where they should have been counted in the census according to census residence rules.

Both the P-sample and the E-sample measure enumeration errors in the census. The E-sample measures gross overcount in the form of erroneous enumerations. The P-sample measures gross undercount in the form of nonmatches. Ideally, the entire census would be searched before a P-sample person was declared to be a nonmatch. Likewise, the entire country would ideally be searched to determine if an E-sample enumeration is a duplicate or is fictitious. Of course, such extensive searches are not feasible operationally. Therefore, the size of the search area must be limited. A limited search will increase the gross overcount and gross undercount. The gross overcount and gross undercount should balance in order to determine net coverage error.

Balancing error occurs when inconsistent treatments are applied to the P-sample and the E-sample. For example, the P-sample search is limited to the sample block, but the E-sample work allows expanding the search area to a surrounding ring of blocks.

The 1980 Post Enumeration Program did not have overlapping search areas. The P-sample was selected from the Current Population Survey and the E-sample was selected from housing units enumerated in the census. The two samples could have had different sized search areas causing them not to balance.

Since 1980, the coverage measurement surveys conducted by the Bureau have had overlapping search areas guaranteeing balancing for the P-sample and E-sample in the search area. Under the overlapping sample design being used, the same blocks are included in the P-sample as in the E-sample. The P-sample search area is , by definition, the proper search area. The E-sample search area is chosen to be consistent with the P-sample search area.

With our current design the search area is the block cluster for most clusters. We search in one ring of surrounding blocks for only a small number of clusters. The current plan is to do targeted extended search for approximately 20 percent of the clusters. We must be certain that the searching in surrounding rings is only conducted for the clusters selected to be included in targeted extended search to guarantee a balanced design.

### 6.7 E-sample Indicators

An E-sample indicator identifies the location of a census enumeration within the search area.

1 = The census enumeration is in the E-sample within the sample block cluster.

2 = The census enumeration is in the sample block cluster, but was not in sample in the large cluster.

3 = The census enumeration is in the first ring of blocks surrounding the sample cluster.

The first ring of blocks surrounding the sample cluster is defined as all blocks touching the sample cluster at one or more points.

## 6.8 Housing Unit Definition and Occupancy Status[33]

Living quarters are either housing units or group quarters. They are usually found in structures intended for residential use but also may be found in structures intended for nonresidential use, like commercial buildings, as well as in tents, vans, shelters for the homeless, dormitories, barracks, old railroad cars, squatters in abandoned buildings, and so forth.

A housing unit is a house, an apartment, a mobile home or trailer, a group of rooms or a single room occupied as a separate living quarters or, if vacant, intended for occupancy as a separate living quarters. Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall. For vacant units, the criteria of separateness and direct access are applied to the intended occupants whenever possible. If that information cannot be obtained, the criteria are applied to the previous occupants.

A housing unit is classified as occupied if it is the usual place of residence of the person or group of individuals living in it at the time of enumeration or if the occupants are only temporarily absent, that is, away on vacation.

- The occupants may be a single family, one person living alone, two or more families living together, or any other group of related or unrelated individuals who share living arrangements.

- Occupied rooms or suites of rooms in hotels, motels, and similar places are classified as housing units only when occupied by permanent residents, that is individuals who consider the hotel as their usual place of residence or have no usual place of residence elsewhere.

- If any of the occupants in a rooming or boarding house live separately from others in the building and have direct access, their quarters are classified as separate housing units.

- Living quarters occupied but not meeting the preceding criteria for a housing unit are defined as group quarters and the individuals living in them are group quarters residents. Living units defined as group quarters without current occupants are excluded from the census universe.

[33] The source of this information is a memorandum from John H. Thompson to Robert W. Marx, dated February 3, 1997, Census 2000 Decision Memorandum No. 8, subject: Revised Housing Unit Definition for Census 2000.

- The living quarters occupied by staff personnel within any group quarters are separate housing units if they satisfy the housing unit criteria of separateness and direct access; otherwise, they are considered group quarters.

A housing unit is vacant if no one is living in it at the time of enumeration, unless its occupants are only temporarily absent.

- Units temporarily occupied at the time of enumeration entirely by individuals who have a usual residence elsewhere are classified as vacant.

- New units not yet occupied are classified as vacant housing units if construction has reached a point where all exterior windows and doors are installed and final usable floors are in place.

- Vacant units are excluded from the housing unit inventory if they are open to the elements; that is the roof, walls, windows, and/or doors no longer protect the interior from the elements. Also excluded are vacant units with a posted sign indicating that they are condemned or they are to be demolished.

- Units that are vacant, fit the definition of a housing unit, but are boarded up are housing units.

- Also excluded from the housing unit inventory are structures or rooms being used entirely for nonresidential purposes, such as a store or an office, or space used for the storage of business supplies or inventory, machinery, or agricultural products. It is a housing unit if household goods are stored in the unit.

There were two changes to the 1990 definition of a housing unit and one change to the vacancy classification. The concept of eating separately has been removed from the definition of a housing unit. Also, the conversion of housing units to group quarters when there are nine or more unrelated people living in the housing unit will not be done for Census 2000.

Vacant rooms in hotels, motels, and similar transient facilities where the building was 75 percent or more occupied by permanent residents were treated as housing units. They will not be housing units in Census 2000.

## 6.9 Census Residence Rules

July 26, 1999

| Residence Rules for Census 2000 | | |
|---|---|---|
| **Household Population** | | |
| **Rule 1.** | Person lives in this household but is temporarily absent on Census Day on a visit, business trip, vacation, or in connection with a job (e.g., bus driver, traveling salesperson, boat operator). This includes foreign nationals whose usual place of residence is in the U.S. and American citizens traveling overseas. | **Count person at:**<br>This household |
| **Rule 2.** | Person has multiple residences and, as of Census Day, travels between one residence and another on a "weekly cycle," a "monthly cycle," a "yearly cycle," or some other cycle (e.g., commuter workers, "snowbirds," and children in joint custody situations). | **Count person at:**<br>The residence where they spend most of time during the week, month, or year, etc. If an individual cannot identify such a place for himself/herself, count him/her at the residence where he/she was on Census Day. (See "Guiding Principles" attached for more information.) |
| **Rule 3.** | Person lives in this household, but is in a general or Veterans Affairs hospital on Census Day. Including newborn babies who have not yet been brought home. | **Count person at:**<br>This household, unless in a psychiatric or chronic disease hospital ward, or a hospital or ward for the mentally retarded, the physically handicapped, or drug/alcohol abuse patients. If so, the person should be counted in the hospital. |
| **Rule 4.** | Person is a member of the U.S. Armed Forces stationed on a nearby military installation or ship but on Census Day is living in this off-base household. | **Count person at:**<br>The off-base household |
| **Rule 5.** | Person is a college student not living in this household during the school year and is only here during break or vacation (See Rules 6 and 25). | **Count person at:**<br>The residence where the person lives while attending college (Usual Home Elsewhere (UHE) not allowed). |
| **Rule 6.** | Person is a college student living in this household during the school year (See Rules 5 and 25). | **Count person at:**<br>This household |
| **Rule 7.** | Person is a student attending school away from home below the college level, such as a boarding school or a Bureau of Indian Affairs boarding school. | **Count person at:**<br>This household |

| | |
|---|---|
| **Rule 8.** Person is an officer or crew member of a merchant vessel and on Census Day is engaged in inland waterway transportation. | **Count person at:**<br>This household |
| **Rule 9.** Person works for and lives in this household and has no other home (e.g., a domestic worker or nanny who "lives in"). | **Count person at:**<br>This household |
| **Rule 10.** Person is staying temporarily in this household on Census Day and has another home. | **Count person at:**<br>DO NOT LIST. (This person will be counted at the other household.) |
| **Rule 11.** On Census Day, person is a citizen of a foreign country who has established a household (or is part of an established household) in the U.S. while working or studying. This includes any family member living with the person. | **Count person at:**<br>This household |
| **Rule 12.** Person is a citizen of a foreign country and on Census Day is living on the premises of an Embassy, Ministry, Legation, Chancellery, or Consulate in the U.S. | **Count person at:**<br>This household, that is, the Embassy, etc. (The person has the right to refuse to provide any or all information.) |
| **Group Quarters Population, UHE Allowed** | |
| **Rule 13.** Person is a member of the U.S. Armed Forces and on Census Day is living on a military installation in the United States, or is living on a military vessel which is assigned to a home port in the United States. | **Count person at:**<br>The residence where the person spends most of his/her time (UHE allowed) [GQ code 601 for military barracks on base; GQ code 602 for transient quarters for temporary residents; GQ code 603 for military ships]. If the person does not claim a UHE, count him/her at the military installation or at the home port of the vessel. |
| **Rule 14.** On Census Day, person is at a camp for temporary workers, such as agricultural or migrant workers; or logging, pipeline, or construction workers. | **Count person at:**<br>The residence where the person spends most of his/her time (UHE allowed) [GQ code 901 for agriculture workers' dormitories on farms; GQ code 902 for other workers' dormitories]. If the person does not claim a UHE, count him/her at the camp. |

| | |
|---|---|
| Rule 15. On Census Day, person is at a hostel, YMCA/YWCA, or transient location, such as a commercial or public campground, racetrack, park, or carnival (See also Rule 16). | **Count person at:**<br>The location where they spend most of their time (UHE allowed) [GQ code 908 for hostels or YMCAs/YWCAs; GQ code 910 for commercial or public campgrounds, racetracks, fairs, or carnivals]. If the person does not claim a UHE, count them at the special place. |
| Rule 16. On Census Day, person is at a recreational camp (i.e., a commercial or public campground). This rule is targeted to persons known as "full-timers" or "good-sams" who live and travel in a recreational vehicle, and the recreational vehicle is their only or usual residence. | **Count person at:**<br>The location where the person spends most of his/her time (UHE allowed). If the person does not claim a UHE, count them at the camp. (Note that if the recreational vehicle is their only or usual residence, it is considered a housing unit (HU) and tabulated as a HU. It is part of GQ enumeration but not part of the GQ population.) |
| Rule 17. On Census Day, person is at a soup kitchen or outreach program (e.g., mobile food van). | **Count person at:**<br>The location where these individuals spend most of their time (UHE allowed) [GQ code 704 for soup kitchens; GQ code 705 for outreach program]. If the person does not claim a UHE, count them at the special place. |
| Rule 18. Person is an officer or crew member of a U.S. flag merchant vessel and on Census Day is docked in a U.S. port or is sailing from one U.S. port to another U.S. port. | **Count person at:**<br>These persons are allowed to claim a UHE [GQ code 900]. If they do not claim a UHE, count them at the merchant vessel. |
| Rule 19. Person is a resident staff member or a member of a special place. For example, a staff member living in a hospital or nursing home, or a member of a religious order living in a monastery or convent. | **Count person at:**<br>These persons are allowed to claim a UHE [GQ code 904 for staff members living in military hospitals; GQ code 905 for staff members living in civilian group quarters; GQ code 906 for religious group quarters]. If they do not claim a UHE, they are counted at the special place. |
| **Group Quarters Population, UHE Not Allowed** | |
| Rule 20. On Census Day, person is under formally authorized, supervised care or custody, in a correctional institution, such as a federal or state prison, local jail or workhouse, federal detention center, or halfway house. | **Count person at:**<br>The special place (UHE not allowed) |

| | | |
|---|---|---|
| **Rule 21.** | On Census Day, person is under formally authorized, supervised care or custody, in a nursing, convalescent, or rest home for the aged and dependent. | **Count person at:**<br>The special place (UHE not allowed) |
| **Rule 22.** | On Census Day, person is under formally authorized, supervised care or custody, in a juvenile institution such as a residential school for delinquents. | **Count person at:**<br>The special place (UHE not allowed) |
| **Rule 23.** | On Census Day, person is under formally authorized, supervised care or custody, in a home, school, hospital, or ward for the physically handicapped, mentally retarded, or mentally ill. | **Count person at:**<br>The special place (UHE not allowed) |
| **Rule 24.** | On Census Day, person is at an emergency shelter, including shelters with sleeping facilities for individuals without a usual residence; shelters for abused women; shelters for runaway, neglected, or homeless children; or shelters for other homeless persons. | **Count person at:**<br>The shelter (UHE not allowed) |
| **Rule 25.** | Person is a college student living in a group quarters (e.g., a dormitory, or sorority or fraternity house) (See Rules 5 and 6). | **Count person at:**<br>The group quarters (UHE not allowed) |
| **Overseas Population** | | |
| **Rule 26.** | Person is a member of the U.S. Armed Forces and on Census Day is stationed on a military vessel which is assigned to a home port in a foreign country. | **Count person at:**<br>DO NOT LIST. (This person will be counted as part of the overseas population.) |
| **Rule 27.** | Person is a member of the U.S. Armed Forces and on Census Day is assigned to a military installation outside the United States. This rule includes family members living with him/her. | **Count person at:**<br>DO NOT LIST. (This person will be counted as part of the overseas population.) |
| **Rule 28.** | Person is an American citizen overseas employed by the U.S. government and on Census Day has a place of duty abroad. This rule includes family members living with him/her. | **Count person at:**<br>DO NOT LIST. (This person will be counted as part of the overseas population.) |

| **DO NOT LIST Population** | | |
|---|---|---|
| **Rule 29.** | Person is an American citizen and on Census Day is working, studying, or living abroad, but not employed by the U.S. government. | **Count person at:**<br>DO NOT LIST |
| **Rule 30.** | Person is a citizen of a foreign country who on Census Day is temporarily traveling or visiting in the U.S. | **Count person at:**<br>DO NOT LIST |
| **Rule 31.** | Person is an officer or crew member of a U.S. flag merchant vessel which on Census Day is docked in a foreign port, is sailing from one foreign port to another foreign port, is sailing from a U.S. port to a foreign port, or is sailing from a foreign port to a U.S. port. | **Count person at:**<br>DO NOT LIST |

## Guiding Principles for the Residence Rules as They Apply to Individual(s) with Multiple Residences

The following provides guidance for determining "usual residence" for an individual with more than one residence.

### Weekly Cycle

If a person is on a "weekly cycle," he/she should be counted at the residence where he/she spends most of their time during the week. For example:

> Some individuals live part of the week at a residence near where they work, and live at their "family home" the rest of the week. We consider these people to be on a "weekly cycle," and they should be counted at the residence where they spend most of their time during the week (e.g., commuter workers).

### Monthly Cycle

If a person is on a "monthly cycle," he/she should be counted at the residence where he/she spends most of his/her time during the month. For example:

> Some children live with one parent for one week out of the month and the other parent the remaining three weeks during the month. We consider these individuals to be on a "monthly cycle" and they should be counted at the residence where they spend most of their time during the month (e.g., children in joint custody situations).

### Yearly Cycle

If a person is on a "yearly cycle," he/she should be counted at the residence where he/she spends most of his/her time during the year. For example:

> 1. Some individuals live in one state during the spring, summer, and fall, but move to a state in a warmer climate during the winter months ("snowbirds"). We consider these people to be on a "yearly cycle," and they should be counted at the residence where they spend most of their time during the year.
>
> 2. Some college students live at the college during the school year and at the "family home" during holidays or the summer. We consider these people to be on a "yearly cycle," and they should be counted at the residence where they spend most of their time during the year.

### No Clearly Defined Cycle

If a person is on no clearly defined "cycle," he/she should be counted at the residence where he/she was on Census Day. For example:

> Temporary workers may establish another residence for an undefined period of time for work. We consider these people to be on an "undefined cycle," and they should be counted at the residence where they were on Census Day.

Time Split Equally Among Two or More Residences

No matter what the cycle, if time is split equally among multiple residences, a person should be counted at the place where he/she was on Census Day.

### 6.10 Census 2000 Type of Enumeration Areas[34]

The Census Bureau defines type of enumeration area (TEA) codes at the census collection block level. Each block must have a TEA code, and no block may have more than one TEA code.

TEA 1 – Block Canvassing and Mailout/Mailback

- Contains areas with predominantly city-style (house number/street name) addresses used for mail delivery
- Census address list is created from United States Postal Service (USPS), 1990 census, local/tribal, and other potential supplementary address sources
- Blocks are included in both Block Canvassing and the Postal Validation Check
- Blocks are included in local/tribal program to identify "new construction"

Mailout/mailback is the most efficient, cost-effective enumeration method in heavily populated areas in which mail is delivered to city-style addresses in virtually all cases (there may be scattered non-city-style mailing addresses in use in these areas). In most instances, a census enumerator visits a residence once - during Block Canvassing. A subsequent visit is sometimes necessary during Nonresponse Follow-up.

The mailing list used for this operation is derived initially from automated address files (the USPS Delivery Sequence File and the 1990 Census Address Control File), and updated through various operations, including Address List Review (LUCA 1998), ongoing DSF updates, Block Canvassing, the Postal Validation Check, and the New Construction Program.

[34] The term type of enumeration area (TEA) has been used for several decennial censuses. For Census 2000, it reflects not only the type of enumeration, but also the method of compiling the census address list that controls the enumeration process.

TEA 2 - Address Listing and Update/Leave

- Contains areas with some number of non-city-style (e.g., P.O. Box or Rural Route) mailing addresses
- Census address list is created from Address Listing, and updated from Address List Review (LUCA) 1999 Recanvassing (in selected areas) and Update/Leave
- Blocks are NOT included in Block Canvassing, the Postal Validation Check, or the New Construction Program
- Puerto Rico, including its military bases, is completely in TEA 2

Address Listing and Update/Leave are implemented in areas where mail often is delivered to non-city-style addresses. In these areas, it is difficult to obtain an up-to-date mailing address list and then "geocode" each address (that is, assign it to a collection block code), because of the constantly changing residential location/mailing address relationship (especially for P.O. Box addresses). The census address list therefore is compiled through a door-to-door independent listing operation (Address Listing) that is implemented in all TEA 2 blocks.

During Address Listing, enumerators knock on each residence door to obtain the occupant's name, phone number, residential address (or location description), and mailing address. (Enumerators do NOT revisit residences whose occupants are not present. This is why the census address list frequently does NOT contain a mailing address, and why the location description is the ONLY "address" in the census address list for many residences.) Enumerators identify the location of each building (containing living quarters) they encounter with a uniquely numbered map spot that they enter on their map and record in their address register; this number is linked to all residential units in the building, and stored in both the census address list and the TIGER data base. These areas will be included in Address List Review (LUCA) 1999.

At census time, enumerators deliver census questionnaires to all housing units compiled during Address Listing and that remain in TEA 2. In the course of delivering these questionnaires, the enumerators also update the census address list and map spotted map to reflect housing units that were not previously listed, and to eliminate residences that they cannot locate. (This operation is called Update/Leave, because the enumerators UPDATE the census address list and maps and LEAVE questionnaires.) Update/Leave enumerators use the residential address/location description in conjunction with the map spot location to determine the correct delivery point for all questionnaires.

Most housing units in TEA 2 areas are visited at least twice by enumerators -- once during Address Listing, and again during Update/Leave. Respondents must mail their completed census questionnaires to the Census Bureau, and so some residences also will be visited a third time, during Nonresponse Follow-up.

TEA 3 - List/Enumerate

- Contains areas that are remote, sparsely populated, or not easily accessible
- Census address list is created and enumeration conducted concurrently
- Blocks are not included in Block Canvassing, the Postal Validation Check, the New Construction Program, or Address Listing
- Includes all military bases in TEA 3 areas
- All Island Areas (except Puerto Rico), including their military bases, are TEA 3

Some areas are remote, sparsely populated, and/or not easily visited. Many of the residences in these areas do not have city-style mail delivery. It is inefficient and expensive to implement Address Listing, Update/Leave, and Nonresponse Follow-up operations involving multiple visits. Instead, the creation of the address list and the delivery/completion of the census questionnaire are accomplished during a single operation, List/Enumerate. Enumerators visit residences in TEA 3 blocks, LIST them for inclusion in the census address list, mark their location on their map with a map spot and number, enter that map spot number in their address register, and ENUMERATE the residents on-site. They collect the same address information as in Address Listing, and include a map spot to reflect each building that contains one or more living quarters. These areas will NOT be included in any Address List Review (LUCA) program, because there is no address list for them in advance of the census.

TEA 4 - Remote Alaska

- Similar to List/Enumerate, but conducted earlier, before ice breakup/snow melt
- These areas will NOT be included in any Address List Review (LUCA) program, because there is no address list for them in advance of the census

TEA 5 - "Rural" Update/Enumerate

- Contains blocks initially in TEA 2, with map spots for all structures containing at least one housing unit
- In some instances, blocks initially in TEA 3 will be converted to TEA 5. These blocks were not included in Address Listing and LUCA 1999, and therefore lack structures and map spots in the MAF and TIGER at the times that LUCA 1999 and "Rural" Update/Enumerate are conducted
- Self-enumeration (through Update/Leave) is thought to be unlikely or problematic
- Census address list is updated, and enumeration is conducted, concurrently
- Blocks are NOT included in the Postal Validation Check or the New Construction Program
- The term "rural" reflects Address Listing as the initial source of the census address list, and does NOT reflect the official census definition of the term "rural"
- These areas will be included in Address List Review (LUCA) 1999 materials, as the MAF was compiled initially from Address Listing

In some areas that otherwise meet the criteria for inclusion in TEA 2, the Census Bureau has decided that having respondents enumerate themselves and return their questionnaires via the mail is not the best way to conduct the enumeration. Some targeted populations may be less likely to return their questionnaires in the mail, and more likely to respond to an enumerator. In other areas, housing units may be vacant because they are occupied seasonally.

In these and comparable situations, enumerators visit all residences on the census address list and complete the enumeration on-site. In the course of delivering these questionnaires, they also update the census address list to 1) reflect housing units that were not previously listed (including a map spot to reflect each building that contains one or more living quarters), and 2) eliminate housing units that they cannot locate. (This operation is called "Rural" Update/Enumerate, because the enumerators work in areas that were Address Listed, UPDATE the census address list [and assign map spots as well], and ENUMERATE the residents.)

TEA 6 - Military

- Contains blocks within TEA 2 that are on military bases
- Mailout/Mailback for family housing
- Separate enumeration procedures for barracks, hospitals, etc.
- Blocks are included in both Block Canvassing and the Postal Validation Check
- These blocks are included in Address List Review (LUCA) 1998 materials, as the MAF was compiled initially in the same manner as TEA 1 areas

The Department of Defense has advised the Census Bureau that virtually all family housing (that is, individual residences as opposed to barracks, hospitals, and jails) are assigned city-style addresses to which the Postal Service delivers mail. The Census Bureau therefore implements mailout/mailback methods to enumerate the population of these individual residences. Within TEA 1 areas, blocks on military bases are assigned a TEA code of 1. Within TEA 2 areas, blocks on military bases are assigned a TEA code of 6. There is no difference between TEA 1 blocks on military bases and TEA 6 blocks in terms of either compiling the census address list or enumerating the population. Blocks within military bases in List/Enumerate areas (TEA 3) also are TEA 3.

TEA 7 - "Urban" Update/Leave

- Contains blocks initially in TEA 1
- Census address list is updated, and questionnaires are delivered concurrently, by Bureau staff (following procedures employed in TEA 2 areas, but without assigning map spots)
- Blocks ARE included in the Postal Validation Check and the New Construction Program
- The term "urban" reflects the predominance of city-style addresses, and does NOT

- reflect official census definition of the term "urban"
- These blocks are included in Address List Review (LUCA) 1998 materials, as the MAF was compiled initially in the same manner as TEA 1 areas

In many areas where mail is delivered mostly to city-style addresses, older apartment buildings are common. In many of these buildings, unit designators (that is, apartment numbers), often do not exist. Further, the subdivision of existing units into multiple units, and the conversion of non-residential space to living quarters, may be frequent. Mail, therefore, often is not delivered to individual apartments (or individual mail boxes), but instead left at common drop points.

In some other areas with mostly city-style addresses, many residents have elected to receive their mail at post office boxes. The Census Bureau is concerned that the city-style addresses of these residents may not appear in the census address list.

To ensure questionnaire delivery to the largest number of residences, update/leave procedures are employed. As these residences have city-style addresses, there is no need for enumerators to assign map spots to assist enumerators in identifying these residences in subsequent operations.

TEA 8 – "Urban" Update/Enumerate

- Contains blocks initially in TEA 1, without map spots for any addresses; maps generated for TEA 8 areas will not include map spots
- Contains mostly blocks on those American Indian Reservations that initially were included in both TEA 1 and either TEA 2 or 3
- Same enumeration procedures as TEA 5
- The term "urban" reflects the initial inclusion of the block in TEA 1 due to the predominance of city-style mailing addresses
- These areas are included in Block Canvassing and the Postal Validation Check

Most American Indian Reservations will be enumerated using a single enumeration procedure (Mailout/Mailback, Update/Leave, or Update/Enumerate). Some of these initially contained blocks with a mixture of TEA codes. In these instances, the reservations will be enumerated using Update/Enumerate methods (see TEA 5). However, for affected blocks initially in TEA 1, the MAF and TIGER do not include map spots for structures containing at least one housing unit. Instead of converting these blocks to TEA 5 ("Rural" Update/Enumerate) and determining map spot locations, the blocks are being distinguished by a separate TEA.

TEA 9 - Additions to Address Listing Universe of Blocks

- Contains groups of blocks (Assignment Areas) initially assigned to TEA 1
- Converted to Address Listing before Block Canvassing is conducted
- Blocks are NOT included in Block Canvassing, the Postal Validation Check, or the New Construction Program

Some blocks that are in TEA 1 contain a significant number of living quarters with non-city-style addresses. These blocks should not be included in Block Canvassing, which is an operation that is designed to confirm and correct the existence and/or location of city-style addresses. The Geography and Field Divisions are identifying Block Canvassing assignment areas (AAs) that likely contain blocks with significant numbers of non-city-style addresses. Some of these AAs will be removed from Block Canvassing, and included in Address Listing. The blocks in these AAs will be assigned a TEA code of 9, and the census address list compilation and census enumeration activities in TEA 9 blocks will be virtually identical to those in TEA 2 blocks (for instance, they will be included in Update/Leave and Nonresponse Follow-up).

Because most of these blocks had few, if any, addresses in the MAF from the USPS, the entities the bloc are in mostly had nothing to review during Address List Review (LUCA) 1998. For this reason, most of these blocks will have their address list reviewed during a new phase of LUCA, often called "LUCA 99

### 6.11 Coverage Improvement Follow-up Universe

The five categories of cases that are eligible for Coverage Improvement Follow-up (CIFU) are:

- Vacant or deleted housing units identified in nonresponse follow-up
- Blank mail return forms
- New construction adds
- Late adds from the update/leave operation or the postal validation check
- lost mail return forms

For more details see DSSD CENSUS 2000 PROCEDURES AND OPERATION MEMORANDUM #CC-2, Definition of the Coverage Improvement Follow-up Universe for Census 2000, dated July 6, 1999.

### 6.12 Census Use Boxes on Short Form Nonresponse Follow-up Forms

The definitions for the census use boxes on the short form for nonresponse follow-up enumerator forms are listed below:

| | | |
|---|---|---|
| A | = | Status on April 1, 2000 |
| B | = | POP on April 1, 2000 |
| C | = | Vacant categories |
| D | = | Interview in Spanish |
| E | = | Usual Home Elsewhere |
| F | = | Mover |
| G | = | Partial Interview |
| H | = | Refused |
| I | = | Replacement |
| J | = | Closeout, unit status and pop count only |

| | | |
|---|---|---|
| K | = | currently not used |
| L | = | JIC1 |
| M | = | JIC2 |
| N | = | JIC3 |
| O | = | JIC4 |

### 6.13 Just In Case

The census questionnaires have items to collect data called just in case (JIC). There are four of these JIC boxes. They are defined as follows:

| Just In Case | Application | Character | Values |
|---|---|---|---|
| JIC1 | All mail returns | First Character | 8 = Transcribed from foreign language form<br>9 = Translated from foreign language form |
| | | Second Character | 0 = English (US)<br>1 = Spanish<br>2 = Chinese<br>3 = Korean<br>4 = Tagalog<br>5 = Vietnamese<br>6 = English (Puerto Rico)<br>7 = Spanish (Puerto Rico) |
| JIC2 | not yet assigned | | |
| JIC3 | All forms | First Character | 1 = Regular GQ enumeration<br>2 = T-Night location |
| | | Second Character | 1 = Respondent filled form<br>2 = Enumerator interview<br>3 = Enumerator filled from administrative records<br>4 = Other |
| JIC4 | All SEQ adds | | Field Operations two digit task code |

### 6.14 Definition II

This section is from the Hogan and Cowan paper titled Imputations, Response Errors, and Matching in Dual System Estimation" published in the 1980 Proceedings of the Section on Survey Research

Methods on pages 263 to 268.

Definition I - A person is "correctly enumerated" if he should have been enumerated and was enumerated once and only once, even though it might have been in an incorrect location. A person is "missed" if he should have been enumerated in the census but was not enumerated in any location. An enumeration is considered to be an "erroneous enumeration" if the person should not have been enumerated but was (e.g., he did not exist, lived outside the U.S., was born after the census or died before the census), or the person should have been enumerated but was enumerated more than once.

Definition II - A person is "correctly enumerated" if he was enumerated in the census at the address reported by the followup survey as the census date residence. A person is "missed" if he was not enumerated at the census date residence that was reported in the followup survey. An enumeration is considered to be "erroneous" if the followup survey reports that the person was not living at the location where the census recorded him. For example, the followup survey could report that no such person exists, or that the person was born after the census, died before the census or was living elsewhere on census date.

The Census Bureau has found that it is impossible to search all locations where a person might have been enumerated. So we are forced into Definition II. But, while seemingly clear for the purpose of defining misses, the definition must be carefully used in dealing with erroneous enumerations. In theory, where one reports one should have been enumerated should be the same regardless of how one is sampled for a followup survey (System 2).[35] But for the people who move between the census and the followup survey, serious problems can arise. This brings us to our next issue: misstatement of address.

One of our serious problems is that many people misstate their Census Day address. Many people report that they were living "here" during the followup survey even though they have moved. A less common problem is people who report their address as "there" during the followup survey even though they moved before Census Day. This phenomenon, known as telescoping, has been uncovered in other studies withe the same net result. Careful probing can reduce this problem, but it cannot eliminate it. Clearly, anyone who misstates their Census Day address will be counted as missed. This must be properly balanced in the E-sample by treating people who misstate their address as erroneously enumerated. There are two ways of doing this: one potentially unbiased, but expensive, one potentially biased but cheap.

The potentially unbiased method is to followup outmover, and interview them at their new address. The interview would be normal "System 2" survey interview. They would be asked "where they were living on Census Day". If they correctly reported their previous address, they would be counted as correctly enumerated there. If they incorrectly reported their old address, we would treat them as erroneously enumerated at the old address. Thus, the treatment of misreporting of address is the

[35] System 2 is A. C. E. interview or the P-sample.

estimation of erroneous enumerations would be consistent with the estimation of omissions.

The other approach is to accept the word of the current occupant as to who was living there on Census Day. Thus, if the current occupant wrongly reports that he was living "here" on Census Day we accept this. If he also reports that the previous occupants moved out before he moved in, we accept that. Clearly, any other family enumerated in the housing unit at the time of the census was erroneously enumerated–if we accept the word of the current occupants! Again, the reports may be inaccurate but they are consistent and balancing.

The methods we have outlined are a way to handle a difficult problem. However, they do not solve the problem, and more than hot-decking has solved the problem of nonresponse.

As always, field work should be done so as to minimize nonresponse, and erroneous enumerations. Matching rules should be constructed to keep the insufficient information category as small as possible. But the problem will exist and all one can do is to attempt to handle it in an unbiased manner.

## 7.0 Glossary

- A. C. E. housing unit - Housing units listed in the A. C. E. independent listing books.

- Preliminary enhanced list - The preliminary enhanced list contains all housing units confirmed to exist in the block cluster. The following housing unit codes are on the preliminary enhanced list: M, MU, UI, UE, CI, and CE. This updated listing is used as the inventory of housing units for subsampling large blocks.

- Subsampled preliminary enhanced list - The subsampled preliminary enhanced list is the listing of housing units after subsampling is completed.

- P-sample housing unit - Housing units listed in the listing book who are confirmed to exist in the block cluster and are not sampled out for A. C. E. Matches to the census housing units and nonmatches that are confirmed to exist within the block cluster at the time of the follow-up interview are included in the P-sample. Census housing unit identified as correctly enumerated or with unresolved enumeration status are not included in the P-sample. Sampling is done within large block clusters.

- Enhanced List - The enhanced list contains all housing units in the P-sample. The P-sample is interviewed to collect inmovers, outmovers, and nonmovers.

- E-sample housing unit - Housing units who are counted in the census when the person matching begins and are not sampled out for A. C. E. Sampling is done within large block clusters.

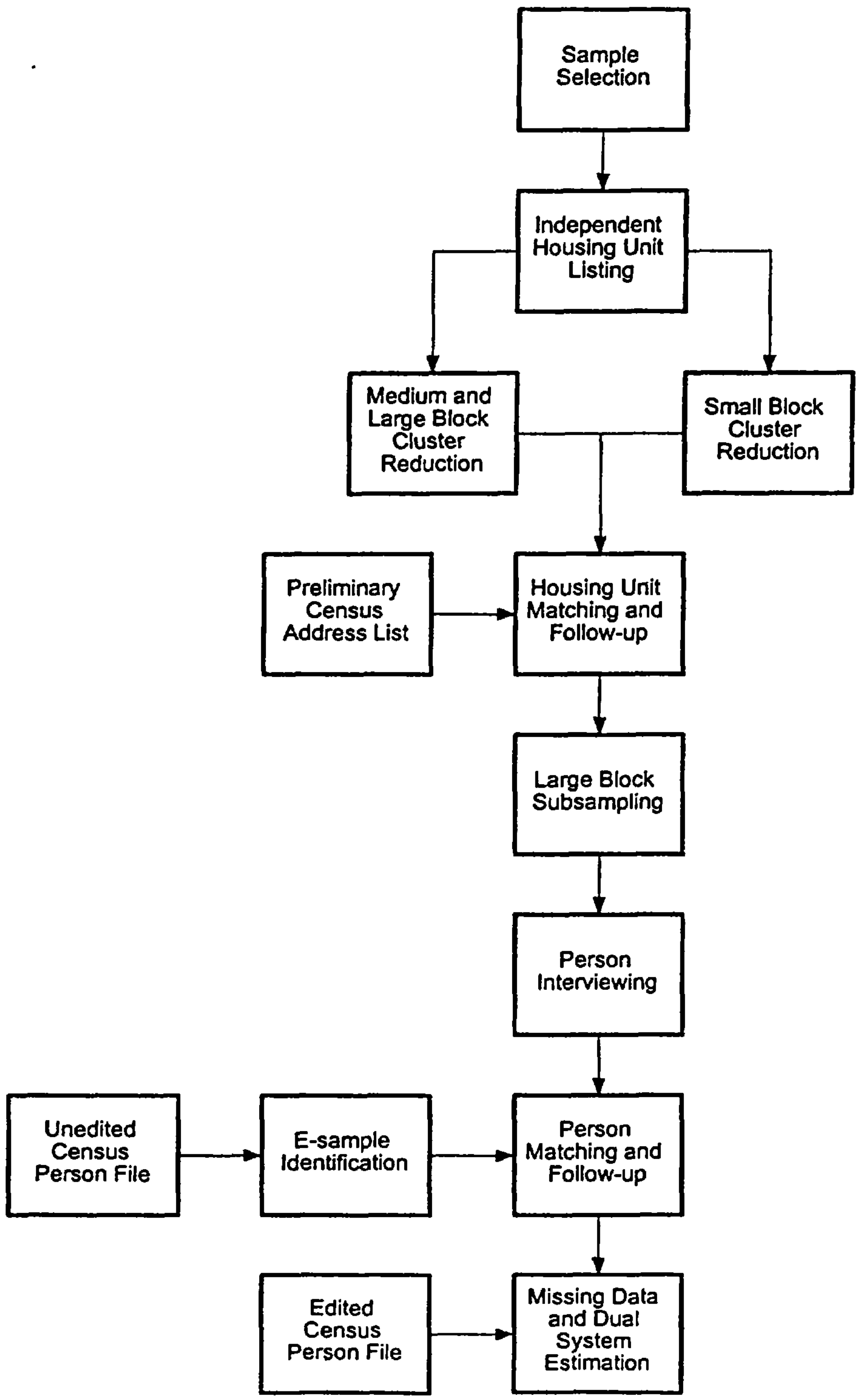
- A. C. E. people - The people collected within the person interview.

- P-sample people - People who were identified as nonmovers or outmovers and were residents of the A. C. E. housing unit on census day. (People identified as nonmovers or outmovers and their residence status is unresolved are included with the P-sample people to attempt to resolve their residence status during the A. C. E. person follow-up operation.) The people must be from an P-sample housing unit.

- E-sample people - The people enumerated in E-sample housing units in the sample block cluster. There are non E-sample census people in large block clusters after subsampling. People counted in the census in group quarters are part of the non E-sample census people. Matching between the P-sample and non E-sample people is permitted, but non E-sample people who do not match are not sent for a follow-up interview.

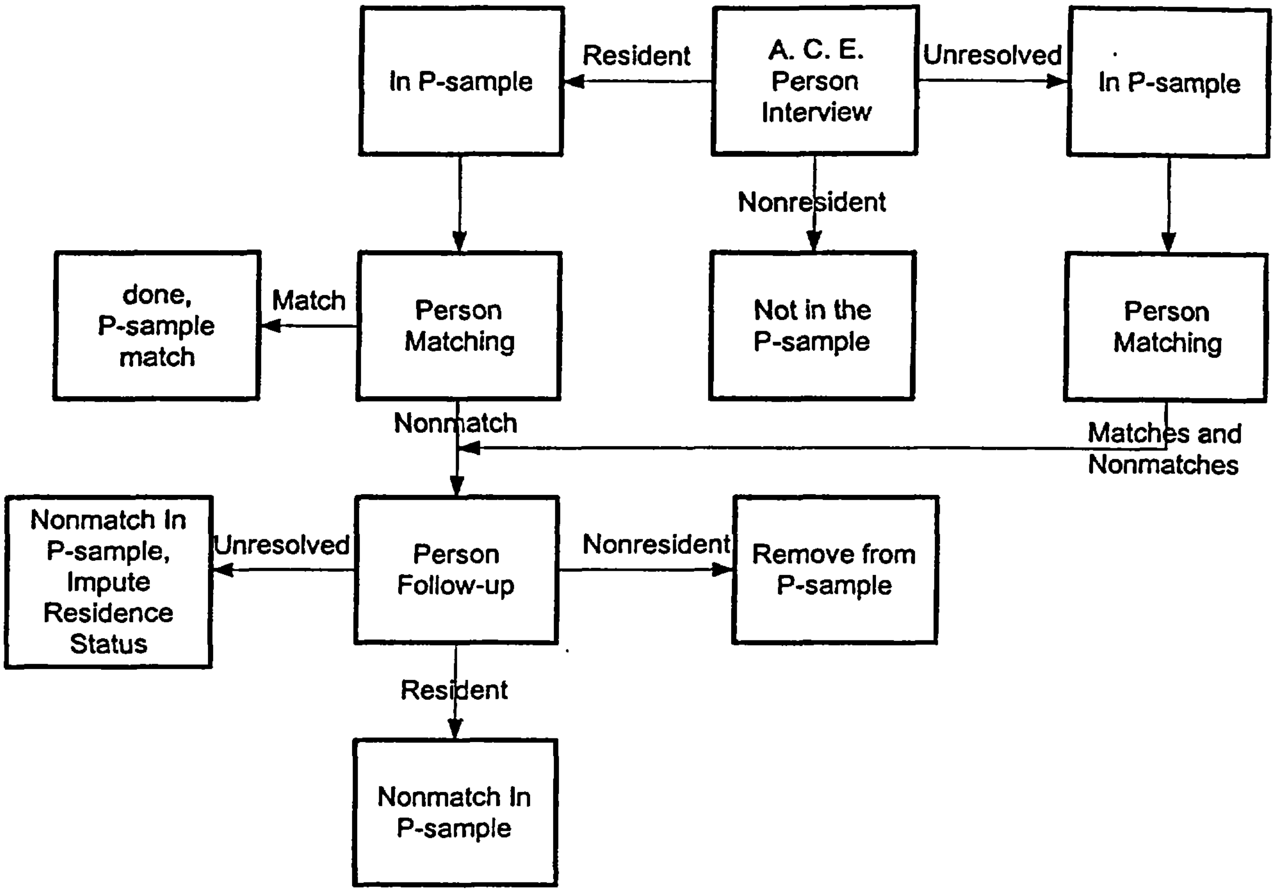
- Roster people - The people in large households on mail returned census questionnaires. There is only a name for person 7 through person 12, if the large household follow-up is not returned. A person record is not created and they are whole person imputations. The names are used to reduce the P-sample nonmatch follow-up workload. The data is only used in partial household nonmatches when there is only one census questionnaire. In other words, the rosters are only used when there is a mail return and no other questionnaires for the household. Person 7 through person 12 in the extended roster on the short form and all people in the household roster on the long form are roster people.

# Attachment 1: Workflow for the Accuracy and Coverage Evaluation

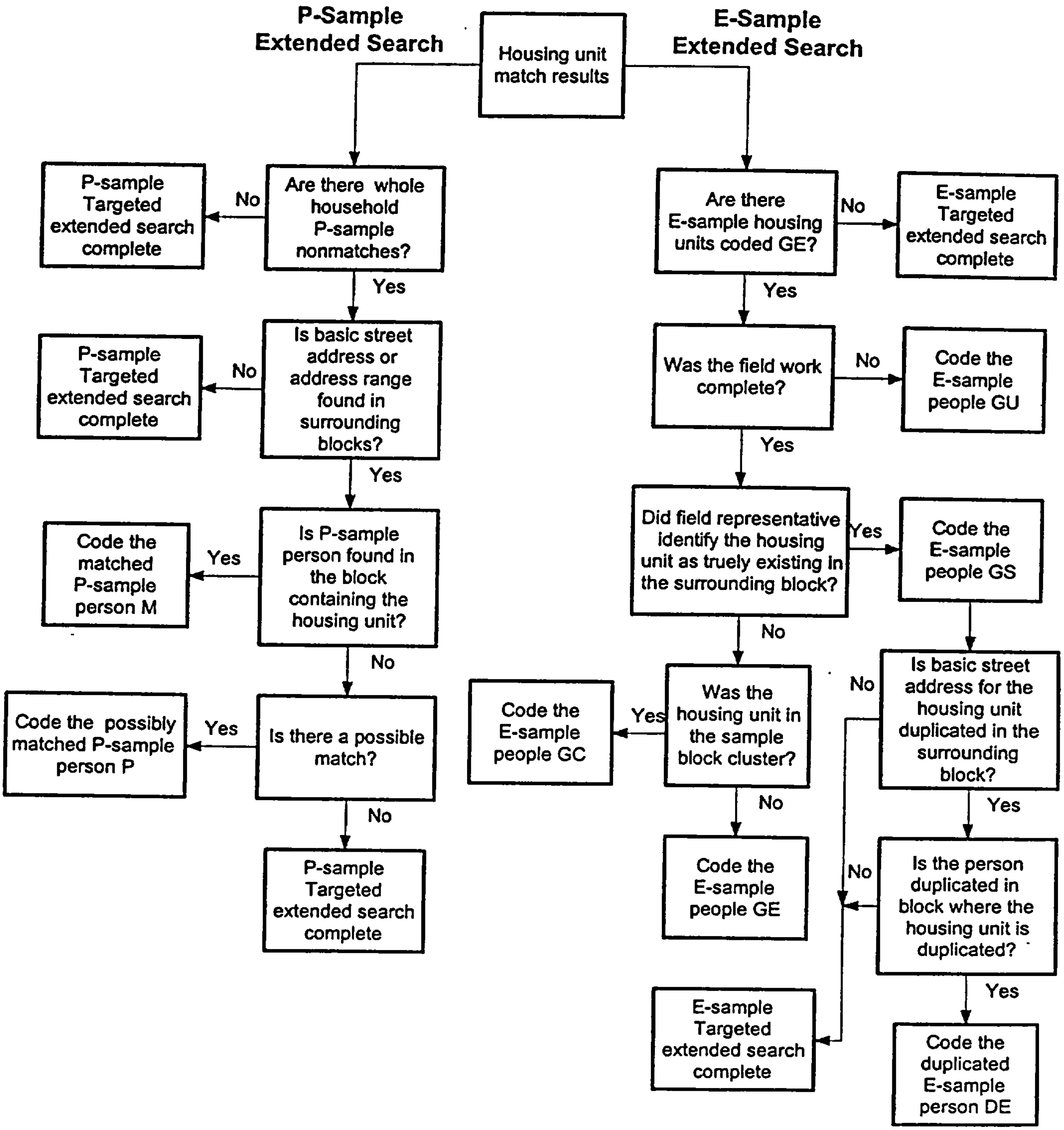


Sample Selection
Independent Housing Unit Listing
Medium and Large Block Cluster Reduction
Small Block Cluster Reduction
Preliminary Census Address List
Housing Unit Matching and Follow-up
Large Block Subsampling
Person Interviewing
Unedited Census Person File
E-sample Identification
Person Matching and Follow-up
Edited Census Person File
Missing Data and Dual System Estimation

# Attachment 2: P-sample Flow



# Attachment 3
# Targeted Extended Search


P-Sample Extended Search
Housing unit match results
E-Sample Extended Search
Are there whole household P-sample nonmatches?
No
P-sample Targeted extended search complete
Yes
Is basic street address or address range found in surrounding blocks?
No
P-sample Targeted extended search complete
Yes
Is P-sample person found in the block containing the housing unit?
Yes
Code the matched P-sample person M
No
Is there a possible match?
Yes
Code the possibly matched P-sample person P
No
P-sample Targeted extended search complete
Are there E-sample housing units coded GE?
No
E-sample Targeted extended search complete
Yes
Was the field work complete?
No
Code the E-sample people GU
Yes
Did field representative identify the housing unit as truely existing in the surrounding block?
Yes
Code the E-sample people GS
No
Was the housing unit in the sample block cluster?
Yes
Code the E-sample people GC
No
Code the E-sample people GE
Is basic street address for the housing unit duplicated in the surrounding block?
No
Yes
Is the person duplicated in block where the housing unit is duplicated?
No
Yes
Code the duplicated E-sample person DE
E-sample Targeted extended search complete

## Attachment 4: The A. C. E. Schedule for 2000

| Activity | Start | Finish |
|---|---|---|
| Produce maps for A. C. E. listing - Wave 1 | 06-15-99 | 08-02-99 |
| Produce maps for A. C. E. listing - Wave 2 and 3 | 07-08-99 | 08-24-99 |
| Conduct listing | 09-07-99 | 12-08-99 |
| Key listing books | 10-04-99 | 01-21-00 |
| Obtain census housing unit inventory | 01-26-00 | 02-15-00 |
| Conduct computer match - housing unit | 01-31-00 | 02-28-00 |
| Conduct before follow-up matching - housing unit | 02-07-00 | 03-10-00 |
| Conduct housing unit follow-up | 02-22-00 | 04-04-00 |
| Conduct after follow-up coding - housing unit | 03-06-00 | 04-18-00 |
| Subsample large blocks and create enhanced list | 03-22-00 | 05-05-00 |
| Create DMAF telephone file | 03-23-00 | 05-12-00 |
| Conduct Census NRFU Enumeration | 04-27-00 | 07-07-00 |
| Conduct A. C. E. person interviews by telephone | 05-08-00 | 06-13-00 |
| Conduct A. C. E. QA interviewing | 05-11-00 | 09-01-00 |
| Conduct A. C. E. person interviews by personal visit | 06-19-00 | 08-18-00 |
| Conduct nonresponse conversion | 07-27-00 | 09-01-00 |
| Conduct data edit | 07-05-00 | 09-15-00 |
| Perform preliminary outcome coding | 09-13-00 | 09-27-00 |
| Obtain CUF | 09-21-00 | 10-10-00 |
| Conduct computer match - person | 10-02-00 | 10-24-00 |
| Conduct before follow-up clerical match - person | 10-13-00 | 11-06-00 |
| Conduct person follow-up interview | 10-23-00 | 11-21-00 |
| Conduct after follow-up coding - person | 11-06-00 | 11-30-00 |
| Send person data files to headquarters | 11-07-00 | 11-30-00 |

| | | |
|---|---|---|
| Conduct computer processing - final housing unit | 03-07-01 | 04-04-01 |
| Conduct before follow-up match - final housing unit | 03-14-01 | 04-13-01 |
| Conduct final housing unit field follow-up | 03-30-01 | 05-11-01 |
| Conduct after follow-up coding - final housing unit | 04-09-01 | 05-22-01 |